

مبادئ الإحصاء والإحتمالات للعلوم الإدارية والتطبيقية

د/ محمد محمد المزاح

رقم المقرر: 928004

2013م 1434هـ



مبادئ الإحصاء والاحتمالات

للعلوم الإدارية والتطبيقية

الدكتور : محمد محمد المزاح

صنعاء

1434هـ- 2013م

التحــكيم العلمي	أ.د.محمد مفرح العيسائي
التحكيم الفني	أ.د. تيسير نجيب الكيلاني
التصميم التعليمي	د. جمال درهم زيد
المراجعة اللغــوية	د. هادي عبد الله شمسان
التصميم الفني	أ. قابوس محمد أحمد عيضة
تصميم الغلاف	أ. عبد السلام عباس النجدي
الإخراج الطباعي	أ. أكرم محمد المطري

الإشراف العام : قسم إنتاج المقررات

الطبعة الثالثة: 1434هـ ـ 2013م

حقوق الطبع والنشر محفوظة لجامعة العلوم والتكنولوجيا، ولا يجوز إنتاج أي جزء من هذه المادة أو تخزينها على أي جهاز، أو نقلها بأي شكل أو وسيلة الكترونية أو ميكانيكية، أو بالنسخ أو التصوير أو التسجيل، أو بأي وسيلة أخرى، إلا بموافقة خطية مسبقة من الجامعة

يطلب هذا الكتاب مباشرة من مركز جامعة العلوم والتكنولوجيا للكتاب الجامعي

Web Site: ust.edu/Centers/ubc

E-mail: ubc@ust.edu

Tel:00967/ 1- 384078

رقم الإيداع (2008ـ81)

مقدمة المقرر

عزيزي الدارس، عزيزتي الدارسة، مرحباً بك إلى هذا المقرر:

لقد أصبح الإسهام في بناء وتطور الحضارة الإنسانية والتقدم العلمي واجباً تفرضه المعرفة وتمليه الظروف، حيث يتركز البحث العلمي في العديد من مجالاته على الأساليب الإحصائية، كأدوات لا غنى عنها في مجال البحث، وتفسير الظواهر، وبناء التوقعات المستقبلية، بهدف الوصول إلى قرارات كفوءة وموثوقة.

وانطلاقاً من فهم الدور الذي يلعبه علم الإحصاء كوسيلة للبحث العلمي أولاً وكعنصر من عناصر الثقافة العامة للفرد ثانياً نضع بين أيدي طلابنا الأعزاء هذا الكتاب ليكون عوناً على فهم مبادئ هذا العلم الضرورية.

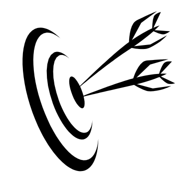
عزيزي الدارس، إن هذا المقرر يهدف إلى تعريفك بدور علم الإحصاء في الحياة وأهميته في الارتقاء بالعملية الإدارية والإنتاجية، كما يعرفك بالطرق المتبعة لجمع البيانات في الدراسات والبحوث العلمية، ويعينك أيضاً على معرفة الفرق بين الإحصاء الوصفي الذي يستعمل في مستويات جمع البيانات، وتنظيمها، وتلخيصها، ووصفها لتكون بصيغة مفهومة وذات مدلول، وبين الإحصاء الاستدلالي الذي يتعلق بطرق تحليل وتفسير وتقدير واستخلاص الاستنتاجات بالاعتماد على الطرق الإحصائية المتقدمة. كما يحتوي هذا المقرر على أساليب عرض البيانات الأولية، في جداول منظمة وترتب فيها البيانات بتوزيعات تكرارية مناسبة، والتوزيعات التكرارية جدولياً وبيانياً، وهذه الطرق البيانية تدربك على عرض البيانات برسوم بيانية تساعدك في فهم، ووصف البيانات فتتعرف على صفاتها وخصائصها من حيث تحديد طبيعة التوزيع من حيث التماثل، والالتواء.

وتتعرف في هذا المقرر على مقاييس النزعة المركزية، ومقاييس التشتت المطلقة والنسبية، ومقاييس التماثل والالتواء، كما يمكنك التعرف على كيفية تحديد العلاقة بين الظواهر المختلفة من خلال عرض موجز لنظرية الارتباط، والانحدار الخطي البسيط، وتوضيح نوع العلاقة بين متغيرين كميين من خلال لوحة الانتشار التي تميز العلاقة الخطية من غير الخطية. وكذا يتناول المقرر دراسة السلاسل الزمنية، والتي يحتاجها معظم المهتمين من مختلف التخصصات والمستويات العلمية.

ونستعرض في هذا المقرر بعض المفاهيم العامة لنظرية الاحتمالات من خلال استعراض مفهوم الاحتمال ومسلماته وقوانينه، ودوره في الحياة اليومية للفرد، كما تناولنا المتغير العشوائي وأقسامه، والتوزيعات الاحتمالية من خلال تعرفنا على واحد من أهم التوزيعات الاحتمالية وهو التوزيع الطبيعي ودالة كثافته الاحتمالية، وكيفية رسم منحناه، وطريقة استخدام جدول التوزيع الطبيعي القياسي لإيجاد المساحات تحت المنحنى.كما يتعرض هذا المقرر لاختبار الفروض الإحصائية والذي يعتبر من المواضيع الرئيسة في مجال اتخاذ القرارات.

الأهداف العامة للمقرر

عزيزي الدارس، أرحب بك مرة أخرى إلى هذا المقرر والذي يتوقع منك بعد دراسة هذا المقرر وتنفيذ جميع الأنشطة والتدريبات الواردة فيه أن تكون قادراً علي:



1- معرفة المفاهيم الأساسية لعلم الإحصاء ودوره في الارتقاء بالعملية الإدارية والإنتاجية.
2- الارتقاء بمستوى الطالب من المرحلة الوصفية إلى مرحلة القدرة على اتخاذ القرار بناء على ما يتوفر لديه من بيانات.
3- إكساب الطالب القدرة على اختيار وتوظيف الأساليب الإحصائية المناسبة التي اكتسبها لمعالجة وحل المشكلات الاقتصادية، والإدارية، والاجتماعية.
4- إكساب الطالب القدرة على تطبيق الأساليب الإحصائية من خلال المعطيات والظواهر الإحصائية تحت الدراسة مجسداً المعنى الوظيفي للإحصاء.
5- لفت انتباه الطلاب إلى أن علم الإحصاء أصبح الأداة التي لا غنى عنها في مجالات البحث، والتفسير، وبناء التوقعات المستقبلية، واتخاذ القرارات.
6- حساب مقاييس النزعة المركزية ومقاييس التشتت، ومعامل الارتباط الخطي البسيط، ومعادلة الانحدار الخطي البسيط وتقدير معالم النموذج بطريقة علمية.

وقد صمم هذا المقرر لتحقيق الأهداف السابقة، وهو يتألف من أربع وحدات دراسية تشتمل على مناقشة تفصيلية لجميع الجوانب الأساسية والخاصة بدراسة مبادئ الإحصاء، والاحتمالات. ولا شك أن الطريقة المثلى لتحقيق الأهداف العامة لهذا المقرر هي دراسة هذا المقرر بكاملة لذا فقد أشتمل هذا المقرر على الوحدات التالية:

الوحدة الأولى: (المناهج الإحصائية وأساليب عرض البيانات): وقد تناولت مقدمة عامة عن علم الإحصاء، وطرق جمع البيانات، العينات العشوائية، وأساليب عرض البيانات الإحصائية الأولية، والتوزيعات التكرارية.

الوحدة الثانية: (المقاييس الإحصائية): وتشمل مقاييس النزعة المركزية، ومقاييس التشتت المطلق والنسبي، ومقاييس التماثل، والالتواء.

الوحدة الثالثة: (الارتباط والانحدار والسلاسل الزمنية) وقد تناولت هذه الوحدة دراسة العلاقة بين متغيرين كميين باستخدام معامل الارتباط الخطي البسيط، أو العلاقة بين متغيرين نوعيين باستخدام معامل ارتباط الرتب. كما استعرضت الوحدة السلاسل الزمنية للاحتياج الشديد لها في كثير من التخصصات.

الوحدة الرابعة: (مبادئ الاحتمالات): استعرضنا في هذه الوحدة، المفاهيم الأساسية لنظرية الاحتمالات، ومسلمات الاحتمال، وقواعده الأساسية، وكذلك مفهوم المتغير العشوائي، والتوزيع الطبيعي، واختبارات الفروض الإحصائية.

كما اشتمل هذا المقرر على العديد من التدريبات، وأسئلة التقويم الذاتي، والوسائل المعينة التي تمنحك الفرصة لمراجعة أجزاء الوحدة، وأيضاً تطبيق المهارات والمعارف التي اكتسبتها في حل ومعالجة المشكلات التي تعترضك في الحياة العملية.

أهمية دراسة المقرر:

تنبع أهمية دراسة هذا المقرر من أهمية الدور الذي يلعبه علم الإحصاء في الحياة السياسية، والاجتماعية، والاقتصادية، والثقافية حيث يُمكِّن القائمين على الأبحاث وفي فروع العلم المختلفة والمهتمين من حل المشكلات واتخاذ القرارات المناسبة.

ونتيجة للتطور العلمي والتقني وتنامي أعداد البيانات بشكل كبير وتنوعها فإننا نرى أن الإحصاء يكتسب أهمية كبيرة في النواحي الإدارية والاقتصادية واهتمامه لم يقتصر على جمع وحصر البيانات بل تعداها إلى تقييم البيانات وتحليلها واتخاذ قرارات سليمة والتنبؤ بالمستقبل. ومن خلال تفحص العناصر الرئيسة للعملية الإدارية نلاحظ بوضوح تام مدى ارتباط الإحصاء بالإدارة لأن التخطيط السليم يعتمد على قدر كاف من البيانات الإحصائية، والقدرة على فهم البيانات الإحصائية وأساليب استخدامها تعتبر ضرورة لمن يشارك في العملية التخطيطية أو يستفيد منها، حيث لا يمكن مراجعة وتقييم الأداء وقياس فاعليته دون استخدام البيانات والأساليب الإحصائية.

كما أن الرقابة وتطوير معايير الأداء ورفع التقارير والتقييم يحتاج لمعرفة الأساليب الإحصائية، وهنا تبرز أهمية اتخاذ القرار المناسب الذي هو من أهم مسؤوليات الإدارة.

ودراسة هذا المقرر سوف تكسب الطالب القدرة على اختيار وتوظيف الأساليب الإحصائية المناسبة التي اكتسبها لمعالجة وحل المشكلات الاقتصادية، والإدارية، والاجتماعية في حياته العملية.

ولا يفوتني هنا أن أتقدم بعظيم شكري وتقديري للأخوين العزيزين الدكتور/علي عبد الجبار كشيح، والأستاذ/ جمال درهم زيد اللذان كان لهما الفضل بعد الله في تشجيعي على استكمال وإخراج هذا الكتاب والشكر موصول لمن كان له طرف جهد أو مساعدة.

وجدير أن نشير إلى أننا سنرحب بكل ملاحظات زملائنا الأساتذة المختصين في هذا الميدان، ومن جميع القراء لهذا الكتاب فشأنه شأن أي كتاب علمي سيأخذ الملاحظات البناءة عسى الاعوجاج يستقيم والخلل يصوب إذ كما يقول المثل: " العين تجهل نفسها وتدرك ما قد حلى في موضع الشهب ".

المؤلف

جامعة العلوم والتكنولوجيا
وقل رب زدني علما

مصادر التعلم:

المرجع الرئيس للمقرر هو كتاب: مبادئ الإحصاء والاحتمالات للعلوم الإدارية والتطبيقية د. محمد محمد المزاح.

عزيزي الدارس، ولمزيد من المعلومات والاستفادة في هذا المجال يمكنك الرجوع إلى المراجع التالية:

1. أبو صالح، محمد صبحي ومروة أحمد.(2005) : مبادئ الإحصاء. الطبعة الثانية، منشورات جامعة القدس المفتوحة، عمان: الأردن.
2. أبو صالح، محمد صبحي وعوض، عدنان محمد.(1995): مقدمة في الإحصاء . الطبعة الثانية، مركز الكتب، لأردني.
3. البلداوي، عبد الحميد .(1997) : الإحصاء للعلوم الإدارية والتطبيقية . الطبعة العربية الأولى، دار الشروق، عمان : الأردن.
4. رمضان، زياد .(1997) : مبادئ الإحصاء الوصفي والتطبيقي والحيوي.الطبعة الرابعة، دار وائل للنشر، عمان : الأردن.
5. العواد، منذر حسين .(2003): مبادئ الإحصاء.الطبعة الأولى، مركز الأمين للنشر، صنعاء : اليمن.
6. المنصوب، محمد عبد الكريم .(1998): مفاهيم أساسية في الإحصاء . الطبعة الأولى، منشورات دار الخبرة، صنعاء: اليمن.
7. Neter, J.; Wassrman , M.H. and Kutner. , Applied Linear Regression Methods. London: Richard D. Irwin, Inc., 1983.

مصادر أخرى للتعلم الذاتي:

1- موقع عمادة التعلم عن بعد **http//:dl.ust.edu** بنظام L.M.S وموقع الجامعة **www.ust.edu.ye** .

2- السيديهات CD المقدمة من مركز التعلم عن بعد والمعدة خصيصاً للمقرر.

3- نماذج الامتحانات وإجاباتها والمعدة من قبل مركز التعلم عن بعد، بوحدة الإشراف الأكاديمي.

محتوى الكتاب :



وقل رب
زدني علماً
جامعة العلوم والتكنولوجيا


UNIVERSITY OF SCIENCE & TECHNOLOGY
وقل رب زدني علماً
جامعة العلوم والتكنولوجيا

1 الوحدة الأولى

المناهج الإحصائية
(جمع وعرض البيانات)


UNIVERSITY OF SCIENCE & TECHNOLOGY
وقل رب
زدني علماً

محتويات الوحدة

1- المقدمــة:

1.1 تمهيد:

عزيزي الدارس، أرحب بك في مطلع هذه الوحدة، وأدعوك لقراءة هذه الوحدة الخاصة بالمناهج الإحصائية (جمع وعرض البيانات الإحصائية) والتي تهتم بتعريفك بعلم الإحصاء بشكل عام ومجال الدراسة في هذا العلم، إذ يهتم كثير من المختصين بدراسة عدد من الموضوعات النظرية وتطبيقاتها في المجالات المختلفة، وتوضح لك **عزيزي الدارس** هذه الوحدة مدى ارتباط هذا العلم بالنواحي الاقتصادية والإدارية والسياسية...الخ.

وتتكون هذه الوحدة من ثلاثة أقسام رئيسة: يدور موضوع القسم الأول حول عدة أمور مهمة تتعلق بوظائف الإحصاء الأساسية وتعريفه وأنواعه مع التركيز على المقارنة بين الإحصاء الوصفي والاستدلالي، والمقاييس الإحصائية ومفهوم المتغير وأنواعه، الذي يعتبر من الجوانب الرئيسة التي تحدد نوع الطريقة الإحصائية المستخدمة في اختبار الفرضيات الإحصائية، ومفهوم المجتمع الإحصائي.

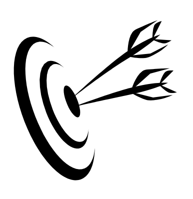
كما تتناول في قسمها الثاني موضوع العينات من حيث تعريف العينة، وطرق اختيار العينات، وأنواع العينات العشوائية.

وخصص قسمها الثالث لدراسة أساليب عرض البيانات الإحصائية جدولياً وبيانياً من خلال عرض البيانات الأولية بواسطة الجداول أو الطريقة البيانية مثل طريقة الأعمدة البيانية، والرسوم الدائرية. كما ستتعلم في هذا القسم طريقة رسم المدرج التكراري، والمضلع التكراري، والمنحنى التكراري، وهذا يساعدك على التدرب على تمثيل الجداول للبيانات التجميعية. ومراعاة للتسلسل المنطقي في استكمال وصف البيانات، نستعرض في نهاية الوحدة التماثل والالتواء والتفرطح في التوزيعات التكرارية، بغرض تعريفك على وصف التوزيعات التكرارية من خلال الرسوم البيانية التي تمثلها، بحيث تستطيع التمييز بين التوزيعات التكرارية المتماثلة وغير المتماثلة.

1 .2 أهداف الوحدة:

عزيزي الدارس، مرحباً بك إلى دراسة الوحدة الدراسية الأولى وهي بعنوان " المناهج الإحصائية وأساليب عرض البيانات" والذي **يتوقع منك بعد دراسة هذه الوحدة أن تكون قادراً علي أن:**

1- توضح الفرق بين الإحصاء الوصفي والإحصاء الاستدلالي.
2- تفرق بين المتغيرات النوعية والمتغيرات الكمية.
3- تناقش مقاييس المتغيرات الإحصائية.
4- تذكر الطرق المختلفة لجمع البيانات الإحصائية .
5- تقارن بين أسلوب المسح الشامل والمسح بالعينة.
6- تبرر استخدام أسلوب المسح بالمعاينة على أسلوب المسح الشامل.
7- تذكر الفرق بين المجتمع الإحصائي والعينة.
8- توضح الفرق بين العينات الاحتمالية ، وغير الاحتمالية.
9- تشرح أنواع المعاينات العشوائية.
10- تعرض البيانات الإحصائية في جداول مناسبة وحسب التصنيفات الملائمة أو المطلوبة وتمثلها بيانياً.
11- تمثل مجموعة من البيانات الإحصائية باستخدام الأعمدة البيانية والرسوم الدائرية.
12- تذكر القواعد التي يج1- توضح الفرق بين الإحصاء الوصفي والإحصاء الاستدلالي.
2- تفرق بين المتغيرات النوعية والمتغيرات الكمية.
3- تناقش مقاييس المتغيرات الإحصائية.
4- تذكر الطرق المختلفة لجمع البيانات الإحصائية .

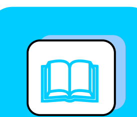


3.1 أقسام الوحدة

عزيزي الدارس الفت انتباهك إلى أن هذه الوحدة تتكون من ثلاثة أقسام رئيسة أعدت لكي تحقق ا لأهداف الأساسية لهذه الوحدة.حيث ارتبط القسم الأول منها (مفاهيم إحصائية عامة) ارتباطاً وثيقاً بالأهداف من (1-5) حيث ارتبط الهدف الأول بمفهوم علم الإحصاء وأنواعه بينما ارتبط الهدفين الثاني والثالث بأنواع المتغيرات الإحصائية والمقاييس الخاصة بها ، وأما مصادر وطرق جمع البيانات فقد ارتبط بالهدف الرابع، وأما أساليب جمع البيانات فقد اتصل اتصالاً وثيقاً بالهدف الخامس . بينما يتصل القسم الثاني والخاص بالعينات، وطرق اختيار العينات العشوائية ، وأنواع المعاينات العشوائية. فيرتبط ارتباطاً وثيقاً بالأهداف من السادس إلى التاسع.

وركز القسم الثالث (عرض البيانات الإحصائية) على الطرق المختلفة لتمثيل البيانات من خلال الرسوم البيانية المختلفة ، والذي يرتبط بالهدفين العاشر والحادي عشر. وترتبط الأهداف من الثاني عشر حتى الخامس عشر بالتوزيعات التكرارية من حيث إنشاء الجدول التكراري والقواعد العامة لإنشاء التوزيع التكراري وطرق التمثيل البياني لتلك الجداول.

1 .4 القراءات المساعدة:



إن المراجع الآتية تمثل قراءات إضافية مساعدة تتعلق بالموضوعات المتضمنة في هذه الوحدة، والذي يرجى منك عزيزي الدارس أن تستفيد منها قدر الإمكان نظراً لاتصالها المباشر بموضوع هذه الوحدة.

1- رمضان، زياد .(1997) : مبادئ الإحصاء الوصفي والتطبيقي والحيوي.الطبعة الرابعة، دار وائل للنشر، عمان : الأردن.

2- المنصوب، محمد عبد الكريم.(1998): مفاهيم أساسية في الإحصاء الطبعة الأولى، منشورات دار الخبرة، صنعاء : اليمن.

3- العواد ، منذر حسين. (2003): مبادئ الإحصاء.الطبعة الأولى، مركز الأمين للنشر، صنعاء : اليمن.

1 .5 الوسائط التعليمية المساندة:



عزيزي الدارس لكي تحقق أهداف هذه الوحدة يجب عليك أن تقوم بالآتي:

- قراءة المادة العلمية الموجودة في هذه الوحدة وحل تدريباتها والتقويم الذاتي الخاص بها.
- عرض شرائح موضح عليها أجزاء من المادة التعليمية.
- جداول الأرقام العشوائية لاختيار العينات.

1 .6 ما تحتاج إليه لدراسة الوحدة:

عزيزي الدارس، نلفت انتباهك إلى أنك أثناء دراستك لهذه الوحدة ستكون بحاجة إلى الآتي:

- دفتر وقلم لكتابة المفاهيم الأساسية التي تناولتها الوحدة. وتحتاج أيضاً إلى قلم رصاص وممحاة ومسطرة وورق رسم بياني.
- أقلام ملونة لتظليل الأشكال المختلفة عند عرض البيانات باستخدام الرسوم البيانية. كما أحثك على تهيئة المكان المناسب للدراسة، ليكون عوناً لك على فهم واستيعاب الأفكار الرئيسة للوحدة من خلال حلك لأسئلة التقويم الذاتي الواردة في ثنايا الوحدة.

2. مفاهيم إحصائية عامة

1.2 النشأة والتطور:

عُرف علم الإحصاء منذ القدم إذ وردت إشارات عن العد من قبل المؤرخ اليوناني (هيرودوتس) حين ذكر أنه في عام 480 ق.م استعمل أحد قادة الجيوش طريقة بدائية بسيطة لمعرفة عدد جيشه.

وفي القران الكريم وردت إشارات كثيرة تقرن الإحصاء بعملية العد قال تعالى : ﴿ لَّقَدۡ أَحۡصَىٰهُمۡ وَعَدَّهُمۡ عَدّٗا ﴾ (مريم /94)، ﴿ ..... وَإِن تَعُدُّواْ نِعۡمَةَ ٱللَّهِ لَا تُحۡصُوهَآۗ ﴾ (إبراهيم / 34) وفي السيرة النبوية الشريفة مثال رائع لعملية العد ذلك عندما قام النبي (ﷺ) بتقدير بالغ الدقة لعدد جيش قريش يوم بدر حينما علم أن قريش تنحر لجيشها كل يوم تسع من الإبل.

ولقد شهدت الحضارات القديمة استخدام أساليب حسابية وإحصائية كثيرة دلت عليها إنجازات تلك الحضارات، ومع ظهور وتطور الدولة ازداد الاهتمام بالبيانات والمعلومات وتصنيفها وتبويبها لتستخدم في تسهيل أمور الدولة والمجتمع.

ويمكن القول إن القرن السادس عشر للميلاد شهد بداية لتطور علم الإحصاء في أوربا حيث أدخل كمادة تعليمية تدرس في الجامعات لأول مرة في ألمانيا عام 1748م، كما ورد في أول نشرة إحصائية صدرت في انجلترا عام 1791 م ومنذ أواخر القرن التاسع عشر أصبح يدرس في معظم الكليات والمعاهد التعليمية .

لقد أصبحت استخدامات علم الإحصاء في العقود الأخيرة تنمو باطراد نتيجة التطورات الكبيرة التي طرأت على حياة الإنسان ونشاطاته بمختلف الميادين العلمية والاقتصادية والاجتماعية والإنسانية، إلى الحد الذي رست فيه طرق الإحصاء كجزء ملازم لمعظم نشاطه اليومي.

إن النمو في استخدامات علم الإحصاء ساعد في إدخال تغييرات جذرية في العملية الإنتاجية والإدارية على مستوى التخطيط لها وتطويرها وقياس النوعية، ومعالجة المشاكل، وأصبح الأداة التي لا غنى عنها في مجال البحث وتفسير الظواهر وبناء التوقعات المستقبلية واتخاذ القرارات.

يتكون مصطلح **Statistics** والذي بناؤه **Stat – is – tics** المشتق من كلمة **State** أي الدولة والذي يعني مجموعة الحقائق الخاصة بشؤون الدولة.

2.2 مفهوم علم الإحصاء Definition of Statistics :

تتزايد أهمية علم الإحصاء يوماً بعد يوم حتى أصبح الأداة التي لا غنى عنها لأي جهة أو برنامج بحثي أو إنساني، و قد تعددت تعريفاته نتيجة لاستخداماته المتعددة. وقبل الشروع في تعريف الإحصاء يكون من الأحسن عزيزي الدارس أن نقف قليلاً لنتعرف على كلمة الإحصاء التي قد تطلق ويراد بها شيئان مختلفان، إذ قد يراد بها تعداد الأشياء أو تصنيفها كقولنا مثلاً.، إن عدد الطلاب في الجامعة 1200 طالب، أو إن نسبة النجاح في الثانوية العامة هي85٪ .

أما المعنى الثاني لكلمة الإحصاء فقد تنصرف إلى ذلك العلم الذي يهتم بتعميم النتائج المستخلصة من العينات على المجتمعات، أو على ما يعرف أيضاً بالاستدلال الإحصائي.

تعريف(1) الإحصاء:

هو علم يدرس كيفية جمع وترتيب وتنظيم وتلخيص وتحليل البيانات، واستخلاص النتائج وتفسيرها، بهدف الوصول إلى نتائج مقبولة وقرارات سليمة.

تعريف(2) الإحصاء:

هو العلم الذي يهتم بتوفير الحقائق الرقمية للظواهر المختلفة بالطرق العملية لجمع وتنظيم وتلخيص وعرض وتحليل البيانات، واستخلاص النتائج، واستعمالها في التنبؤ للتحقق من بعض الظواهر.

ولعل أبسط هذه التعريفات هو الذي يرى أن علم الإحصاء هو **علم اتخاذ القرارات في ضل عدم التأكد.**

3.2 أقسام علم الإحصاء :

1. الإحصاء الوصفي Descriptive Statistics:

يختص بطرق جمع البيانات ووصفها لتكون بصيغة مفهومة وذات مدلول دون الوصول إلى استنتاجات أو استدلالات خاصة بالبيات. أي هو التعامل مع البيانات الإحصائية دون التعميم .

والغرض من الإحصاء الوصفي هو: تقدير معالم المجتمع الإحصائي ووصفه تمهيداً للوصول إلى استنتاجات عنه، فهو عادة خطوة تسبق الإحصاء الاستدلالي.

2- الإحصاء الاستدلالي Inferential Statistics

يتعلق بطرق تحليل وتفسير وتقدير واستخلاص الاستنتاجات، بالاعتماد على جزء (عينة) من المجتمع الأصلي للتوصل إلى قرارات تخص المجتمع الإحصائي وعليه فالإحصاء الاستدلالي هو الذي يتعامل مع التعميم والتنبؤ والتقدير .أي أنه يهتم بتطبيق الأساليب الإحصائية في المجالات المختلفة مثل الاقتصاد والطب والزراعة والعلوم.أي اتخاذ أفضل قرار ممكن عندما تكون المعلومات المتوفرة غير وافية.

تدريب (1)



ما الفرق بين الإحصاء الوصفي والإحصاء الاستدلالي؟

4.2 البيانات الإحصائية Statistical Data

البيانات: هي مجموعة الحقائق والمعلومات التي تتعلق بظاهرة ما وتشكل المادة الخام التي يعتمد عليها علم الإحصاء.

المعلومات: هي بيانات تمت معالجتها بوسائل إحصائية فتحولت إلى شكل مفيد له فائدة وذا معنى

وتنقسم البيانات الإحصائية إلى قسمين: **البيانات الكمية والبيانات النوعية "الوصفية".**

المتغير: هو خاصية قابلة للتغير من مشاهدة لأخرى في المجتمع الإحصائي مثل الدخل، والجنس، والعمر، ونوع العمل، والحالة الاجتماعية، والحالة التعليمية وغيرها.

وتصنف المتغيرات الإحصائية إلى نوعين:

1. المتغيرات الكمية Quantity Variables

وهي المتغيرات التي تأخذ قيماً عددية، لذا يطلق عليها اسم: المتغيرات العددية مثل: عدد طلبة كلية الاقتصاد، وعدد العاملين في الجامعة، وعدد حوادث الطرق السنوية في اليمن، والدخل، والعمر، والوزن.



وتنقسم المتغيرات الكمية إلى:

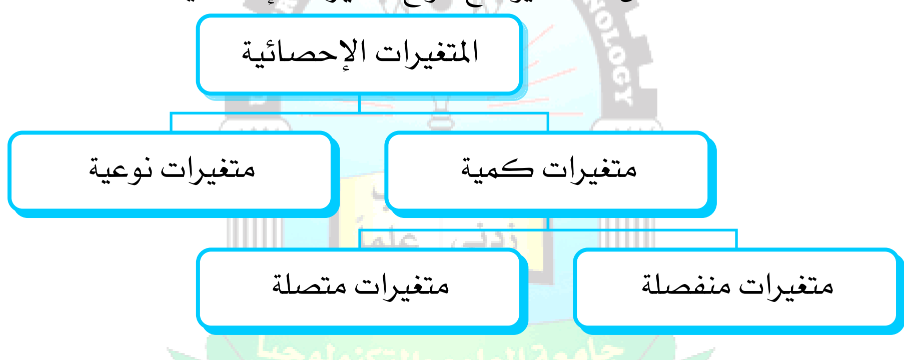
أ- المتغيرات المنفصلة: وهي المتغيرات التي تأخذ قيماً عددية صحيحة ولا تحتوي على قيم كسرية، أي لا توجد قيمة بين قيمتين مثل: عدد طلبة كلية الاقتصاد، عدد العاملين في الجامعة، عدد حوادث الطرق السنوية في اليمن.

ب- المتغيرات المتصلة: وهي المتغيرات التي يمكن أن تأخذ أية قيمة عددية في مدى معين مثل الدخل السنوي أو الشهري، العمر، الوزن.

2.المتغيرات النوعية (المصنفة) أو الوصفية Qualitative Data

وهي تلك المتغيرات التي تصنف المشاهدات في عدة مجموعات تشترك كل مجموعة في صفة معينة مثل الجنس (ذكر، أنثى)، نوع العمل (عامل، موظف، عمل حر) والحالة الاجتماعية (أعزب، متزوج، أرمل، مطلق) والحالة التعليمية(أمي، أساسي، ثانوي، جامعي) والشكل (1.1) يوضح أنواع المتغيرات الإحصائية:

شكل (1.1) يوضح أنواع المتغيرات الإحصائية



5.2 مقاييس المتغيرات الإحصائية:

تنقسم المقاييس الإحصائية حسب نوع المتغيرات إلى قسمين أساسيين والموضحة بالشكل (2.1) وهما:

مقاييس المتغيرات النوعية إلى قسمين:

1. مقياس اسمي: يكون المتغير مقاساً بمقياس اسمي إذا كان لا يمكن وضع أصنافه(أقسامه) في ترتيب معين (درجات). وتكون الأصناف في المتغيرات الاسمية عبارة عن أسماء لذا يطلق على هذا النوع من المقاييس مقياس اسمي مثل: الحالة الاجتماعية التي يمكن تصنيفها إلى (أعزب، متزوج، أرمل، مطلق)

ونوع العمل الذي يمكن تصنيفه إلى(عامل، موظف، عمل حر) والجنس الذي يمكن تصنيفه إلى(ذكر، أنثى).

2. مقياس ترتيبي : يكون المتغير مقاساً بمقياس ترتيبي إذا أمكن وضع أصنافه (أقسامه) في ترتيب معين . مثل مستوى الدخل الذي يمكن تصنيفه إلى (عالي، متوسط، منخفض) وتقدير معدل الطالب الذي يمكن تصنيفه إلى (ممتاز، جيد جداً، جيد، متوسط، مقبول، راسب).

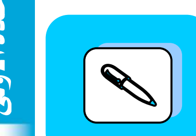
3. مقاييس المتغيرات الكمية وتنقسم إلى قسمين:

أ- مقياس فئوي : يكون المتغير مقاساً بمقياس فئوي عندما تكون المسافات بين النقاط في المقياس متساوية، حيث يمكن في المتغيرات الفئوية مقارنة الفرق بين القيم، ولا يمكن المقارنة بين نسب القيم مثل : درجات الحرارة المئوية.

ب- مقياس نسبي : يكون المتغير مقاساً بمقياس نسبي عندما يمكن المقارنة بين نسب القيم، حيث يمكن في المتغيرات النسبية قسمة قيمها الرقمية بحيث تعطي نسباً ذات معنى مثل: الطول ، الوزن.

شكل (2.1) يوضح المقاييس الإحصائية





تدريب (2)

حدد نوع المتغير والمقياس الإحصائي المناسب لما يلي:

مستوى الدخل ، والوزن ، ونوع العمل ، وحوادث الطرق اليومية .

6. 2 مصادر البيانات الإحصائية Sources of Data:

1. المصادر التاريخية أو الوثائقية His torical Sources:

وتشمل البيانات التي يكون مصدرها السجلات والوثائق والميزانيات المالية، التي تتوفر كحصيلة لنشاط الدوائر والشركات المختلفة من خلال ممارسة نشاطها اليومي وما تقوم بإصداره من نشرات وتقارير وتكون على نوعين:

أ. مصادر أصلية: ويقصد بها الجهات التي تقوم بنفسها بجمع البيانات وتهيئتها ونشرها كما هو الحال مع مكاتب الإحصاء والتعداد المركزية.

ب. **مصادر ثانوية: وهي الجهات التي تعتمد في جمع البيانات على جهات أخرى** إلا أنها تقوم بطبعها ونشرها بعد استلامها من المصادر الأصلية كما هو الحال مع المنظمات الدولية والإقليمية التي تقوم بنشر المطبوعات الإحصائية التي تستلمها من الدول الأعضاء.

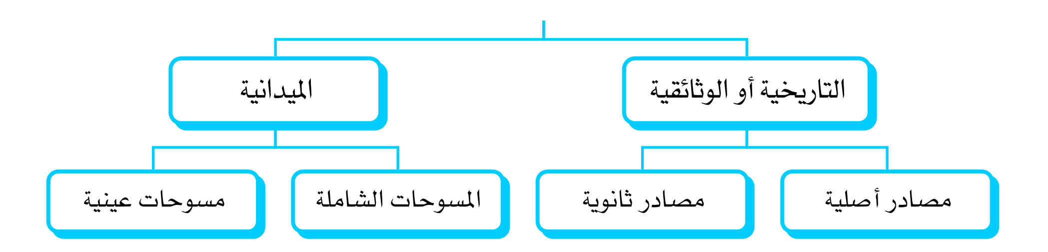
2. المصادر الميدانية Field Sources

وتخص البيانات التي يتم جمعها من وحدات المجتمع الإحصائي ميدانيا بصورة مباشرة بواسطة استمارات وجداول إحصائية تعد لهذا الغرض وهناك أسلوبان لجمع البيانات الميدانية هما:

أ. أسلوب المسوحات الشاملة: الذي يشمل كافة وحدات المجتمع الإحصائي .

ب.أسلوب مسوحات العينة: التي بموجبها يتم جمع بيانات من جزء (عينة)من المجتمع الإحصائي، بطريقة تكون ممثلة لخصائص المجتمع والشكل(3.1) يوضح تلك المصادر.

شكل(3.1) يوضح مصادر البيانات الإحصائية



7.2 طرق جمع البيانات الإحصائية Method of Data Collection

تعددت طرق جمع البيانات الإحصائية بسبب تعدد طبيعة المجتمعات الإحصائية واختلاف البيانات التي نود جمعها والإمكانات المالية المتاحة للدراسة ومنها الطرق الآتية:

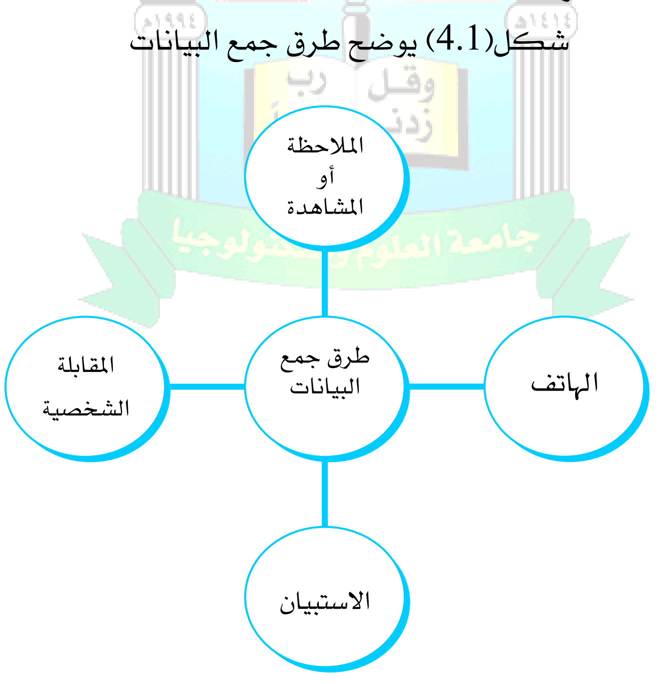
أ. **طريقة المشاهدة أو الملاحظة**: وهي الطريقة التي يتم استخدامها عند مراقبة الظواهر كما هي على الطبيعة، مثل معايشة الباحث لمجتمع السجناء أ و مجتمع البادية.

ب. .**طريقة الاستبيان**: وتعني قيام المبحوثين بتدوين إجاباتهم على الأسئلة الواردة في الاستمارة بأنفسهم.

ج. **طريقة المقابلة الشخصية**: وهي الطريقة التي يتم بواسطتها جمع البيانات من خلال اتصال الباحثين بالمبحوثين شخصياً لأخذ إجاباتهم.

د. **طريقة الهاتف**: وهي أقل أهمية من الطرق السابقة وتستخدم في الحالات التي تنتشر فيها الهاتف بصورة عالية في المجتمع المشمول. و الشكل رقم (4.1) يوضح الطرق الأنفة الذكر.

شكل(4.1) يوضح طرق جمع البيانات



8.2 المجتمع الإحصائي Statistical Population

المجتمع الإحصائي: هو جميع المفردات التي تتمتع بصفة ما أو خاصية ما مشتركة. وهذه المفردات قد تكون بشراً أو أشياء أو ظواهر، وهذا ما يميز المجتمع الإحصائي من المجتمع بالمعنى الاجتماعي، الذي هو مجموعة من الناس تربط بينهم علاقة ما بقصد جمع البيانات العددية عن المجتمع. ولا بد من تحديد المجتمع الإحصائي تحديداً دقيقاً وواضحاً من غير لبس أو غموض حتى نحصل على البيانات الضرورية فقط، ويتحدد المجتمع الإحصائي بتحديد المكان والزمان معا و فيما عدا ذلك لا يتحدد المجتمع الإحصائي، وعلى هذا نرى أن مفهوم المجتمع الإحصائي مفهوماً مرناً حيث **يمكن أن يوسع المجتمع أو العكس بتوسيع المكان أو الزمان أو كليهما** وهذه أمثلة على المجتمع الإحصائي : اسر مدينة صنعاء عام 2005 وعمال إحدى الشركات عام 2004 وطلاب جامعة العلوم والتكنولوجيا عام 2006، ونرى من هذه الأمثلة أن هناك تحديداً للمجتمع من حيث الزمان والمكان.

قد يكون المجتمع الإحصائي محدوداً و ذلك عندما نستطيع حصر عدد مفرداته وبالتالي نستطيع جمع بيانات عنه مثل طلاب جامعة إب سكان مدينة صنعاء.

وقد يكون المجتمع الإحصائي غير محدد وذلك عندما لا نستطيع حصر عدد مفرداته مثل أسماك البحر الأبيض المتوسط أو الاحتياطي من النفط، وهذه المجتمعات يطلق عليها تسمية مجتمعات نظرية.

9.2 أساليب جمع البيانات :

بعد أن يحدد الباحث مصدر البيانات لبحثه يجب أن يحدد الأسلوب الواجب استخدامه لجمع البيانات. وتجمع البيانات بأحد أسلوبين هما : أسلوب الحصر الشامل وأسلوب العينة، واعتماداً على مزايا ومساوئ كل أسلوب وأخذ ظروف البحث بعين الاعتبار يختار الباحث الأسلوب المناسب، لذلك سنتناول أدناه كلا الأسلوبين:

1- أسلوب الحصر الشامل Survey :

يقصد به جمع المعلومات عن الظاهرة موضوع الدراسة من جميع مفردات الظاهرة دون استثناء. ويستخدم هذا الأسلوب عادة في التعدادات العامة مثل تعداد

جامعة العلوم والتكنولوجيا

السكان. ومن مزايا هذا الأسلوب: دقة النتائج، حيث تعطي صورة مفصلة عن جميع مفردات الظاهرة في المجتمع الإحصائي كما هو فعلا باستثناء بعض الأخطاء الحسابية، ولهذا الأسلوب بعض المساوئ منها: أنه يحتاج لوقت طويل، وتكاليف عالية، وتستغرق جهداً كبيراً خاصة عندما يكون حجم المجتمع كبيراً. وقد يتعذر استعمال هذا الأسلوب وذلك عندما يتسبب فحص الوحدات في إتلافها وهذا ما يسمى بالفحص المدمر.

يتعذر استعمال أسلوب الحصر الشامل عندما يتسبب فحص الوحدات في إتلافها وهذا ما يعرف بالفحص المدمر.

2- أسلوب العينة: Sampling Method

العينة يقصد بها ذلك الجزء من المجتمع الذي يتم اختياره عشوائيا أو بصورة غير عشوائية لتمثيل ذلك المجتمع، أي أنها عملية تعميم من الجزء إلى الكل، وعلى أساس تمثيلها لخصائص المجتمع كافة المسحوبة منه العينة فخريجي أحد التخصصات في سنة معينة هم عينة من خريجي نفس التخصص، ومرضى السكري في مدينة ما يمكن اعتبارهم عينة من مرضى السكري في الدولة وهكذا . ونظراً للطبيعة الجزئية لهذه الطريقة فإن نتائجها لا تكون متطابقة تماماً مع نتائج المسح الشامل، كما أنها لا تعطي نتائج مفصلة لمفردات الظاهرة موضوع الدراسة.



تدريب (3)

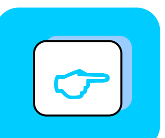
من خلال دراستك للمادة التعليمية في هذا القسم.

اشرح المقصود بالمفاهيم التالية مع ضرب مثال لكل واحد منها:

1- مجتمع محدد 2- الحصر الشامل

3- أسلوب العينة 4- الاستبيان

نشاط:



قم بزيارة للجهاز المركزي للإحصاء واطلع على أساليب إجراء المسوحات الميدانية : ثم صمم استمارة لجمع بيانات تخص القوى العاملة في اليمن.

أسئلة التقويم الذاتي (1) :

?

أ. أكمل الجمل الآتية بما يناسبها :

1- المتغير هو ....

2- الحالة الاجتماعية للفرد هي متغير..............................

3- مستوى الدخل هو متغير......بينما الدخل السنوي هو متغير....

4- الحالة التعليمية للفرد متغير ....ويقاس بمقياس ....

5- الوزن هو متغير : .... ويقاس بمقياس................

6- تتقسم المصادر الميدانية لجمع البيانات إلى ....

7- من عيوب أسلوب الحصر الشامل ....و........و.......

ب. هل تعتبر العينات الآتية ممثلة للمجتمعات المسحوبة منه ؟علل.

1- أرادت إحدى الجرائد التنبؤ بالفائز في انتخابات الرئاسة القادمة بالجمهورية اليمنية فاختارت عينة من 1000 من المشتركين بالجريدة لمعرفة رأيهم.

2- اختيار طلبة الإحصاء لاستطلاع رأيهم بالنسبة لموضوع معين وذلك كعينة من جميع طلبة الجامعة.

جامعة العلوم والتكنولوجيا

3. العينات العشوائية Random Samples

1.3 مفهوم العينة

العينة هي جزء من المجتمع، ويتم اختيارها لتمثل ذلك المجتمع الذي اختيرت منه بإحدى طرق المعاينة.وقد عُرف استخدام العينات منذ القدم حيث يؤدي الغرض المطلوب في كثير من الأحيان بنجاح دون الحاجة إلى أسلوب الحصر الشامل لأنه غالباً ما تكون دراسة المجتمع كاملاً إما مستحيلة أو غير ضرورية، ويفضل أسلوب المسح بالعينة للأسباب الآتية:

إطار المعاينة: هو الذي يوفر التفاصيل الأساسية لوحدات المجتمع الإحصائي قيد الدراسة. مثل سجل الناخبين الذي يحوي أسماء كل من يحق

1- قلة التكاليف: إن معظم الدراسات الإحصائية تكون مقيدة بمقدار من التكاليف والزمن والجهد المخصص لإنجازها لهذا قد لا تسمح هذه القيود بإجراء المسح الشامل.

2. زيادة سرعة إنجاز الدراسة: قد يحتاج الباحث إلى نتائج سريعة أحياناً، ينبني عليها اتخاذ قرار معين الأمر الذي يمكن تحقيقه عن طريق المسح بالعينة بينما يحتاج لوقت طويل باستخدام المسح الشامل، فعلى سبيل المثال استطلاع الرأي العام قبيل الانتخابات أو بعد حادثة معينة.

3- دقة البيانات، قد يحتاج إجراء المسح الشامل إلى عدد كبير من الأشخاص لجمع البيانات وهذا ينتج عنه أخطاء متعددة في البيانات أهمها الفروق الفردية بين العاملين في جمع البيانات، ومن هنا فإن اعتماد المسح بالعينة على عدد أقل من جامعي البيانات المختصين يؤدي إلى تقليل تلك الأخطاء وبالتالي تكون النتائج أكثر دقة.

4- يغطي المسح بالعينة عدد محدد من مفردات المجتمع الإحصائي مما ينتج عنه توفير من الوقت والجهد والموارد.

5- صعوبة المواصلات إذا كان المجتمع متباعد جغرافياً، حيث يتعذر الوصول إلى جميع عناصر المجتمع الإحصائي.

6- قد تكون عناصر المجتمع الإحصائي غير قابلة للعد، وبالتالي يتعذر إجراء المسح الشامل لها، مثل مجتمع الطيور أو دراسة مخزون اليمن من البترول، لذا يفضل المسح بالعينة على المسح الشامل.

لهذه الأسباب وغيرها يفضل استخدام أسلوب العينات لجزء من المجتمع، بشرط أن تكون العينة ممثلة تمثيلاً جيداً لخصائص المجتمع الذي أخذت منه.

إن أسلوب اختيار العينة العشوائية يسمى بالمعاينة Sampling حيث تمثل العينة مجموعة وحدات المعاينة Sampling Units والتي يتم اختيارها من إطار المعاينة وقد تحوي أكثر من وحدة أي أن وحدة المعاينة تضم أشخاصاً أو أسراً أو وحدات إنتاج أي أن وحدة المعاينة قد تتكون من عنصر أو أكثر.

وهناك ثلاثة أمور أساسية يجب مراعاتها عند إتباع أسلوب العينة لجمع البيانات الإحصائية:

1- تحديد المجتمع الإحصائي وإطار المعاينة

2- تحديد المتغيرات التي تجمع منها البيانات في العينة.

3- تحديد نوع المعاينة المستخدمة في الدراسة.

وفي هذا الصدد يجب التمييز بين نوعين من العينات هما:

أ- العينات الاحتمالية أو العشوائية

ب- العينات غير الاحتمالية (التحكمية)

يقصد بالعينات غير الاحتمالية تلك التي يتم اختيارها وفق معايير يحددها الباحث ويعتقد أنها ستؤدي إلى الحصول على عينة ممثلة للمجتمع.

أما العينات العشوائية (الاحتمالية) فهي **عينات يتم اختيارها بأسلوب يسمح بتحديد احتمال ظهور كل مفردة من مفردات المجتمع في العينة ويكون ذلك الاحتمال لا يساوي الصفر.** ويمكن استخدام نتائجها لأغراض الاستدلال الإحصائي.

2.3 طرق اختيار العينات العشوائية

Methods of Selecting Random Samples

سوف نتناول طريقتان لاختيار العينات العشوائية هما:

أ- **طريقة القرعة** (Lottery Method)

تستخدم هذه الطريقة لاختيار عينة عشوائية من مجتمع عدد مفرداته قليل نسبياً حيث تعطى لكل وحدة رقماً متسلسلاً وتكتب هذه الأرقام على بطاقات متماثلة من جميع النواحي بحيث لا يمكن أن تميز إحداها عن الأخريات ثم يتم خلط هذه

البطاقات خلطاً جيداً وتسحب البطاقة تلو الأخرى ويدون ما عليها من رقم إلى أن يكتمل العدد المطلوب اختياره في العينة.

ب- **جدول الأرقام العشوائية** (Random Numbers)

تصلح هذه الطريقة لسحب عينات عشوائية من المجتمعات الصغيرة والكبيرة معاً، لأن طريقة القرعة غير عملية إذا زادت وحدات المجتمع الإحصائي قيد الدراسة عن عدد معين أو حد معين. ولهذا السبب فإن جداول الأرقام العشوائية تم إنشاؤها لكي تستخدم بدلاً من طريقة القرعة. والأرقام العشوائية هي عبارة عن مجموعة الأرقام من 9 , 8 , 7 , 6 , 5 , 4 , 3 , 2 , 1 , 0 خلطت بطريقة معينة جعل من ظهورها في تلك الجداول بنفس تكرار الحدوث، بل إن أي رقم فيها لا يمت بعلاقة لأي رقم آخر على يمينه أو يساره أو إلى أعلاه أو أسفله في الجدول. أي أنها تنتشر في الجدول بشكل عشوائي ويتم سحب العينة على النحو الآتي:

1- تعطى كل وحدة معاينة عدداً متسلسلاً من 1 إلى N، حيث N عدد وحدات المعاينة في الإطار وتكون هذه الأرقام مكونة من نفس عدد المنازل كما في N.

2- اختيار جزء من جدول الأرقام العشوائية تكون أعداده مكونة من نفس عدد المنازل كما في N.

3- تحديد أول عدد يتم قراءته من الجدول بطريقة عشوائية ومن ثم الاستمرار في قراءة الأرقام المتتالية في نظام هندسي ثابت عمودياً أو أفقياً أو قطرياً أي السير بأسلوب منتظم. ولنفرض أننا بدأنا السير من بداية العمود الرابع من أعلى رأسياً فإننا نأخذ العدد أسفل العدد الأول على أساس التتابع الرأسي إلى أن نصل إلى نهاية العمود ثم ننتقل إلى العمود التالي وهكذا حتى نحصل على العدد الذي يساوي حجم العينة المطلوبة. مع ملاحظة أنه:

1- إذا حدث تكرار لأي رقم يتم إلغاؤه واستبداله برقم أخر يليه.

2- يكون التدرج أفقياً أو عمودياً أو قطرياً.



تدريب (4)

ما الفرق بين المجتمع الإحصائي وإطار المعاينة .

وضح ذلك بمثال.

أسئلة التقويم الذاتي (2) :

لماذا يفضل استخدام أسلوب المسح بالعينة عن أسلوب المسح الشامل؟

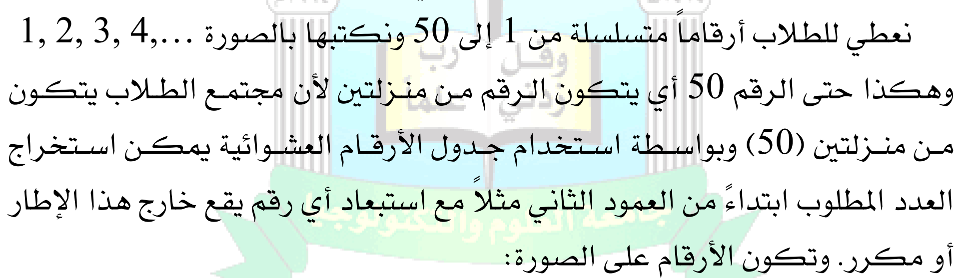
3.3 أنواع المعاينات العشوائية

أ- المعاينة العشوائية البسيطة Simple Random Sampling SRS

هي عينة يتم اختيارها بطريقة تسمح بإعطاء أية عينة ذات حجم n (من مجتمع حجمه N)، وبحجم معين، نفس الفرصة (الاحتمال) لتختار كعينة مدروسة. وبالتالي تكون كل عينة من العينات الممكنة ممثلة للمجتمع ولا يوجد ما يدعو لتفضيل إحداها على الأخريات



مثال(1.1):



إذا كان لدينا مجتمع مكون من 50 طالباً ورغبنا في اختيار عينة عشوائية بسيطة مقدارها 15 طالباً فإنه يجري سحب العينة بواسطة جدول الأرقام العشوائية كما يلي:

نعطي للطلاب أرقاماً متسلسلة من 1 إلى 50 ونكتبها بالصورة 1, 2, 3, 4,... وهكذا حتى الرقم 50 أي يتكون الرقم من منزلتين لأن مجتمع الطلاب يتكون من منزلتين (50) وبواسطة استخدام جدول الأرقام العشوائية يمكن استخراج العدد المطلوب ابتداءً من العمود الثاني مثلاً مع استبعاد أي رقم يقع خارج هذا الإطار أو مكرر. وتكون الأرقام على الصورة:

8 , 35 , 20 , 6 , 1 , 2 , 34 , 33 ,
22 , 27 , 23 , 14 , 25 , 37 , 3

ب- المعاينة العشوائية المنتظمة (Systematic Sampling)

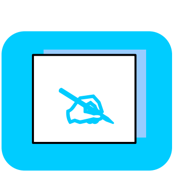
هي بديل آخر للمعاينة العشوائية البسيطة ويتم اختيارها بتقسيم مفردات المجتمع المرتبة والمرقمة من $1, 2 , 3 , ... , N$ إلى فئات متساوية عددها يساوي حجم العينة ونطلق على كل فئة اسم فترة السحب وعدد مفرداتها S ، ونحصل على طول فترة السحب بتقسيم حجم المجتمع N على حجم العينة n أي أن :

$$S = \frac{N}{n}$$

ثم نختار إحدى مفردات الفئة الأولى بشكل عشوائي من مفردات أول فترة سحب ولتكن المفردة Z أول مفردة في العينة ونحصل على باقي المفردات للعينة بإضافة فترة السحب S إلى المفردة الأولى وتكون العينة على الصورة: Z , Z+S , Z+2S , Z+3S , ..., Z+(n-1)S

مثال (2.1):

مجتمع حجم مفرداته 200 مفردة نريد سحب عينة عشوائية منتظمة بحجم 10 مفردات



الحل:

$$\because S = \frac{N}{n}$$

$$= \frac{200}{10} = 20$$

نختار من ضمن فترة السحب الأول والبالغ عددها 20 مفردة بشكل عشوائي بإحدى طرق السحب المذكور سابقاً ولتكن هذه المفردة ذات الرقم 5 فتكون مفردات العينة هي:

5 , 25 , 45 , 65 , 85
105 , 125 , 145 , 165 , 185

ومن مزاياها قلة التكاليف وسهولة السحب كما تعطي نفس النسب الموجودة في المجتمع لصفة ما . ومن عيوبها أنه لا يمكن سحبها إلا إذا كان المجتمع محدود ومفرداته مرتبة ومرقمة كما تتصف بالتحيز في بعض الأحيان .

ج-المعاينة العشوائية الطبقية : (Stratified Random Sampling)

هي إحدى البدائل للمعاينة العشوائية البسيطة ويمكن استخدامها لزيادة دقة النتائج وضمان تمثيل المجموعات المميزة في المجتمع وتستخدم عندما تكون مفردات المجتمع غير متجانسة فنقوم بتصنيف مفردات المجتمع إلى مجموعات متجانسة تدعى كل مجموعة طبقة، ثم نختار عينة عشوائية مستقلة من كل طبقة وتعد جميع هذه العينات الجزئية الناتجة عينة واحدة تسمى بالعينة الطبقية ويعتمد حجم كل عينة جزئية على الحجم النسبي للطبقة. فإذا كان حجم المجتمع N يتألف من4 طبقات مثلاً عدد مفرداتها n_1 , n_2 , n_3 , n_4 حيث أن

$$N = n_1 + n_2 + n_3 + n_4$$

تتميز العينة الطبقية بأنها تمدنا بمعلومات عن:
- العينات الجزئية (الطبقات).
- المجتمع ككل من خلال العينة الكلية.

إلا أن تكاليفها أعلى من العشوائية البسيطة لنفس الحجم.

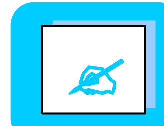
و n حجم العينة الطبقية التي يجب سحبها فنقوم بسحب أربع عينات عينة من كل طبقة أحجامها n_1 , n_2 , n_3 , n_4 فيكون

$$N = n_1 + n_2 + n_3 + n_4$$

وللحصول على قيم n_1 , n_2 , n_3 , n_4 نطبق نسب الطبقات في المجتمع حسب العلاقة:

$$\frac{n}{N} = \frac{n_1}{N_1} = \frac{n_2}{N_2} = \frac{n_3}{N_3} = \frac{n_4}{N_4}$$

مثال (3.1):



أراد باحث سحب عينه من طلاب جامعة العلوم والتكنولوجيا بحجم 80 طالب لمعرفة رأيهم حول النظام الفصلي والساعات فإذا علم أن عدد الطلاب في الجامعة 8000 طالب منهم 4000 طالب يؤيدون النظام الفصلي و3000 يؤيدون نظام الساعات والباقي يؤيدون النظامين فما هو عدد الطلاب الذي يجب على الباحث أخذه من كل طبقة .

الحل:

$$\frac{n}{N} = \frac{n_1}{N_1} = \frac{n_2}{N_2} = \frac{n_3}{N_3}$$

$$\frac{80}{8000} = \frac{n_1}{4000} = \frac{n_2}{3000} = \frac{n_3}{1000}$$

$$n_1 = \frac{320000}{8000} = 40$$

$$n_2 = \frac{240000}{8000} = 30$$

$$n_3 = \frac{80000}{8000} = 10$$

وتمتاز المعاينة الطبقية بأن نتائجها تكون أكثر دقة من نتائج المعاينة البسيطة وذلك لقرب خصائصها من خصائص المجتمع .

د- المعاينة العشوائية العنقودية (Cluster Random Sampling)

إن إتباع أي من الأساليب السابقة يتطلب توفر إطار للمعاينة وقد نواجه في كثير من الحالات بعدم توفر إطار المعاينة ويصبح من المتعذر اختيار عينة بسيطة أو طبقية أو منتظمة والبديل هو السعي لإعداد إطار أو أن نلجأ إلى أسلوب آخر وهو تقسيم المجتمع إلى مجموعات جزئية غير متداخلة تسمى الواحدة منها عنقود بحيث يحتوي

كل عنقود على عدد من مفردات المجتمع المتجاورة مكاناً أو زماناً لاختيار عينة عشوائية من بين تلك العناقيد. وتمتاز المعاينة العنقودية على العشوائية البسيطة بأنها:

1- أقل تكلفة 2- أسهل في التنفيذ 3- تتلافى مشكلة إعداد إطار المعاينة لجميع مفردات المجتمع.

مثال(4.1):



أراد رئيس الدوريات الخارجية معرفة معدل السرعة للسيارات الذاهبة من صنعاء إلى تعز وللقيام بهذه المهمة قامت دورية بتسجيل سرعة كل سابع سيارة تمر من نقطة معينة المطلوب:



1. ما هو مجتمع الدراسة؟ 2- ما هي العينة في هذه الدراسة؟
3. كيف يمكن للدورية أن تعمل مسحاً شاملاً ؟
4. ما طريقة المعاينة التي استعملتها الدورية ؟
5. إذا كان عدد السيارات التي قامت الدورية بتسجيل سرعتها 100 سيارة فكم كان حجم المجتمع ؟

الحل:

1. مجتمع الدراسة هو جميع السيارات التي استعملت الطريق من صنعاء إلى تعز في ذلك اليوم.
2. العينة منتظمة وتكونت من السيارات التي كان رقمها المتسلسل عند مرورها من النقطة 7,14,21 ,28........ الخ
3. يمكن للدوريه أن تعمل مسحا شاملا وذلك بتسجيل سرعة كل سيارة لدى مرورها من النقطة المعينة.
4. طريقة المعاينة التي استعملتها الدورية هي طريقة المعاينة العشوائية المنتظمة.
5. حجم العينة 100 وهذا يساوي $\frac{1}{7}$ حجم المجتمع. إذاً حجم المجتمع يساوي 700 سيارة.



تدريب (5)

نود سحب عينة من طلاب الجامعة لمعرفة رأيهم حول دراسة وتدريس اللغة الأجنبية ومن المعروف مسبقاً أن هناك اختلافاً في صعوبة وممارسة هذه اللغات. فإذا كان عدد الطلاب هو 16000طالب منهم 9000 طالب لغتهم الإنجليزية و6000 طالب لغتهم الفرنسية والباقي لغات مختلفة. حدد نوع المعاينة المناسبة للاختيار. وإذا علم أن حجم العينة المطلوبة هي 120 طالب فما هو عدد الطلاب من كل فئة.

أسئلة التقويم الذاتي (3) :

1- اذكر السبب فيما يأتي:

أ- يفضل غالباً استخدام أسلوب المسح بالعينة على أسلوب المسح الشامل.

ب- يفضل أحياناً استخدام أسلوب المعاينة العنقودية عن بقية أنواع المعاينات الأخرى.

2- أذكر أنواع العينات ؟

3- قارن بين المعاينة الطبقية والمعاينة العنقودية.

4. عرض البيانات الإحصائية (PRESNETATION OF STATISTIC& DATA)

البيانات الإحصائية هي نتاج لتعداد سكاني أو مسح ميداني أو تجربة طبية أو فيزيائية أو كيميائية أو ما شابه ذلك. وعادة ما تأتي هذه البيانات بصورة بيانات خام. فلابد أولاً من أن ترتب وتنظم وتلخص وتعرض بطريقة يسهل معها التعرف على خصائصها ومميزاتها. وهذه الخطوة تعد حجر الأساس والخطوة الأولى في التحليل الإحصائي.

وتستخدم أساليب عرض البيانات الإحصائية بشكل واسع في جميع المجالات والعلوم، فلا يخلو بحث أو كتاب أو صحيفة أو نشرة أخبار تلفزيونية أو إعلان تسويقي من أحد هذه الأساليب.

ولقد اكتسبت أساليب عرض البيانات أهمية كبيرة في الوقت الحاضر لقدرة الحاسبات الإلكترونية على القيام بالحسابات الإحصائية اللازمة ووضع أساليب العرض المختلفة.

وفي حالات كثيرة يكون هدف الباحث من عرض الأرقام اجتذاب انتباه القارئ ليس إلى الأرقام بحد ذاتها وإنما إلى اتجاه هذه الأرقام بالزيادة أو النقصان أو العلاقة بين المتغيرات التي يدرسها أو المقارنة بين مجموعتين أو أكثر من البيانات. وقد يكون هدف الباحث هو تعريف القارئ على الوضع الإجمالي إذ قد لا يهتم كثير من القراء بالتفاصيل الدقيقة فيلجأ الباحث إلى إبراز الصورة الإجمالية للبيانات بواسطة عرضها بالرسوم البيانية المعبرة.

وتعرض البيانات الإحصائية عرضاً جدولياً وبيانياً ولكل منهما أساليب عرض. وفي هذا القسم سنتعرف على أهمية جدولة وتنظيم البيانات وأساليب تلك الجدولة والعرض البياني الأكثر شيوعاً واستخداماً.

يهدف العرض البياني إلى:

1- إبراز الصورة الإجمالية للبيانات.
2- جذب انتباه القارئ إلى اتجاه الأرقام وليس إلى الأرقام بحد ذاتها.
3- إبراز العلاقة بين المتغيرات تحت الدراسة.
4- المقارنة بين مجموعتين أو أكثر من البيانات.

1.4 عرض البيانات الأولية (غير المبوبة) :

في حياتنا العملية توجد كميات كبيرة من البيانات، منها ما هو خاص بالشركات والمؤسسات والوزارات، ومنها ما هو خاص بالتجارب الكيميائية والطبية والزراعية وغيرها. وعرض البيانات بطرق مقالية يكون من الصعب استيعابها والمقارنة بين مفرداتها، لذا كان من الضروري عرض البيانات الإحصائية بطرق شيقة وسهلة الفهم، ومن هذه الطرق:

1.1.4 طريقة الجداول:

وهي عبارة عن وضع البيانات الإحصائية في جداول، وكثيراً ما تستعمل في عرض التغيير الحاصل لظاهرة ما مع الزمن، كالبلدان أو المدارس وغيرها.

وهناك قواعد عامة يجب مراعاتها عند تصميم أي جدول لضمان توفر العناصر الأساسية فيه وهي:

1- أن يكون للجدول عنوان كامل واضح ومختصر ومحدد يعطي فكرة دقيقة وواضحة عما يحويه الجدول من معلومات.

2- أن تكون للأعمدة والصفوف عناوين مختصرة ولكنها غير غامضة.

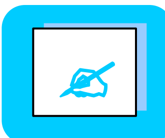
3- أن ترتب البيانات وفق تسلسلها الزمني أو حسب أهميتها من الناحية الوصفية

4- يجب توضيح وحدات القياس المستخدمة

5- توضيح المصدر الذي أخذت منه البيانات الواردة في الجدول

6- ترقيم الأسطر والأعمدة لتسهيل العودة إليها إذا كان لا بد من العودة إليها عند شرح محتوياتها.

7- يجب وضع رقم للجدول إذا كان سيوضع ضمن كتاب وذلك للعودة إليه عند شرح البيانات الموجودة ضمنه.



مثال(5.1):

كان عدد الزوار لأحد المراكز التجارية خلال أسبوع حسب جدول (1.1)



اليوم	عدد الزوار
السبت	850
الأحد	678
الاثنين	645
الثلاثاء	920
الأربعاء	489
الخميس	756

2.1.4 التمثيل البياني للبيانات غير المبوبة:

1.2.1.4 الأعمدة البيانية (Bar Charts)

تعد الأعمدة البيانية من أكثر الطرق شيوعاً في تمثيل البيانات، وتتلخص هذه الطريقة برسم مستطيلات (أعمدة) متساوية القواعد لكل منه اسم على المحور الأفقي، ولكن ارتفاع كل منها يتناسب مع القيمة التي يمثلها ذلك العمود. ويراعى أن يترك بين كل عمود وآخر مسافة مناسبة لفصلهما عن بعضهما. وتوجد عدة أشكال للأعمدة البيانية منها:

أ- **الأعمدة البيانية الأحادية (البسيطة): تستخدم لعرض بيانات تخص ظاهرة واحدة لفترات زمنية مختلفة؛ تهدف إلى ملاحظة التطور التاريخي للظاهرة مع الزمن**. كما في الشكل البياني (5.1) الذي يوضح عدد الطلبة الخريجين من إحدى الجامعات خلال الفترة الزمنية من 2002إلى2007م والموضحة في الجدول(2.1).

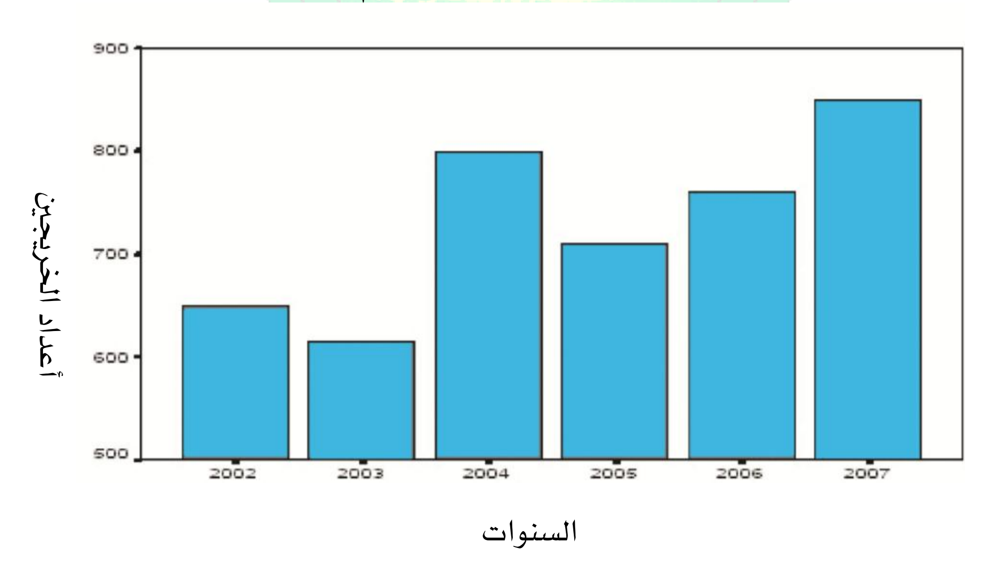
جدول(2.1)

أعداد الطلاب الخريجين خلال الفترة من 1992-1997 من إحدى الجامعات

السنة	2002	2003	2004	2005	2006	2007
عدد الخريجين	650	615	800	710	760	850

شكل (5.1)

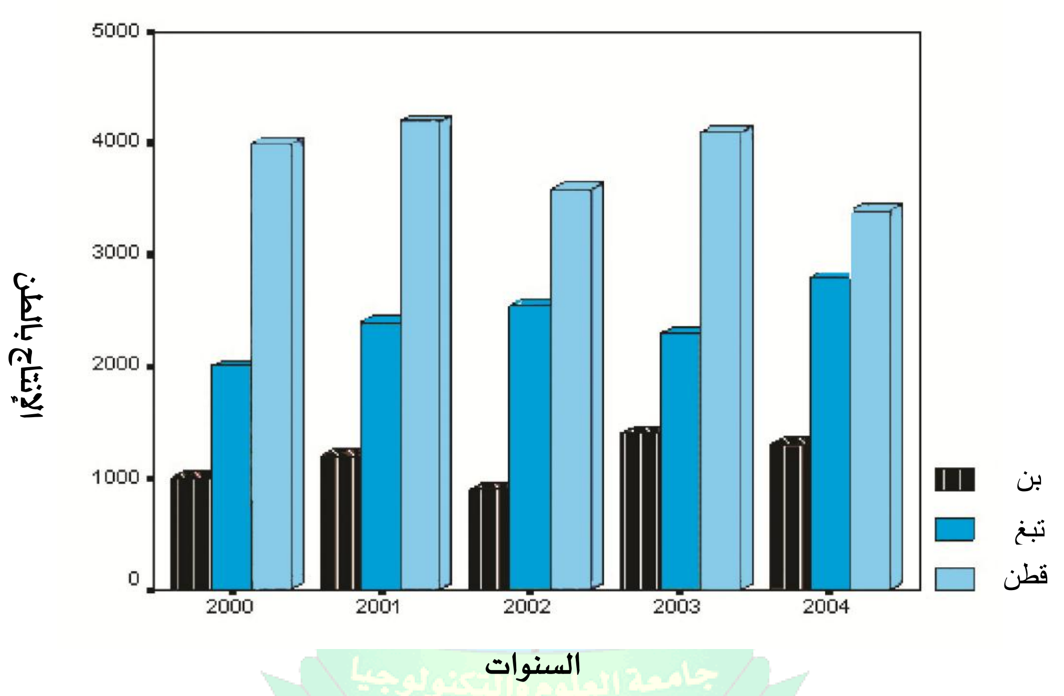
الأعمدة البيانية الأحادية لعدد الخريجين للأعوام 2002 حتى 2007



ب- الأعمدة البيانية المتلاصقة (المتجاورة): تستخدم لعرض ظاهرتين أو أكثر على شكل واحد للفترة الزمنية نفسها بأعمدة متلاصقة يمثل كل منها ظاهرة معينة تميز بلون مختلف مثل مقارنة إنتاج القطن والتبغ والبن للأعوام من 2000 إلى 2004م .

شكل (6.1)

الأعمدة البيانية المتلاصقة لإنتاج القطن والتبغ والبن للأعوام من2000 حتى 2004م

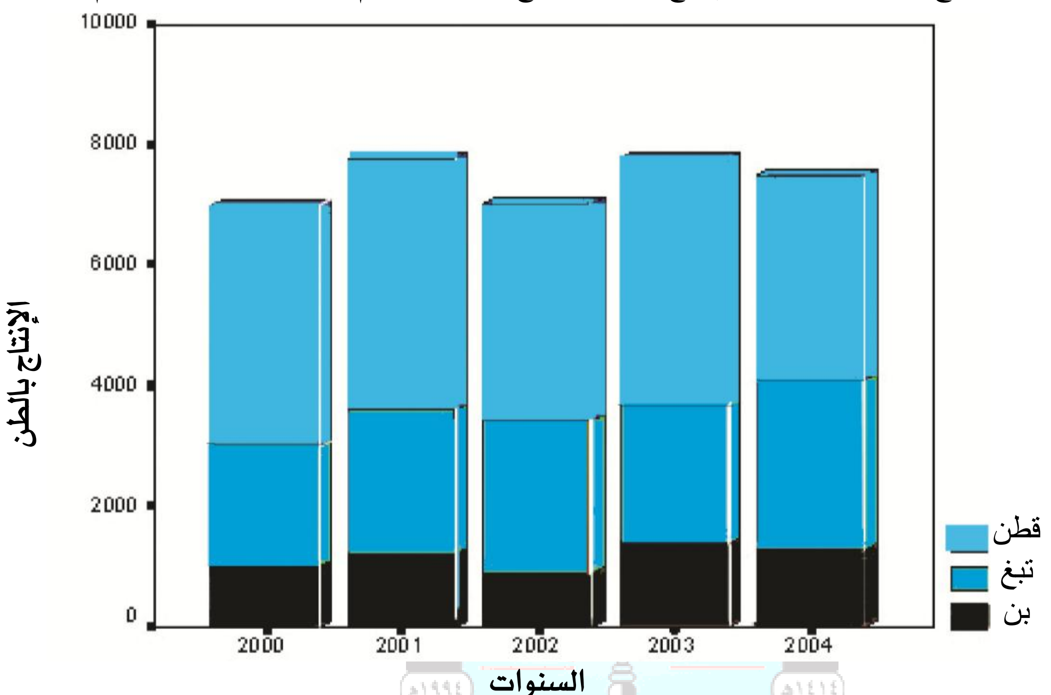


جامعة العلوم والتكنولوجيا

ج- الأعمدة البيانية المركبة: تستخدم لعرض عدة ظواهر أو عدة مستويات لظاهرة واحدة في عمود واحد لنفس الفترة الزمنية بحيث يمثل ارتفاع العمود مجموع قيم الظواهر أو مجموع المستويات للظاهرة المعنية، كما هو موضح في الشكل رقم (7.1) الذي يوضح إنتاج القطن والتبغ والبن للأعوام من 2000 إلى 2004.

شكل (7.1)

يوضح الأعمدة المركبة لإنتاج القطن والتبغ والبن للأعوام من 2000 إلى 2004م



ويمكن أن تأخذ الأعمدة الاتجاه الجانبي، أي أن المحور الأفقي يخصص للكميات أو الأعداد، والمحور العمودي، للصفات أو الأعوام مثلاً ويطلق عليها بالمستطيلات الأفقية.

2.2.1.4 الرسوم الدائرية Pie Charts

الرسوم الدائرية تمثل شكلاً آخر من أشكال العرض البياني والذي يكثر استخدامه في المجال الاقتصادي خاصة، وفيه نعبر عن الظواهر بمساحات دائرية. وتستخدم عندما يكون الهدف ليس متابعة تطور التغيرات التي تطرأ على ظاهرة معينة بل لإبراز الأجزاء التي تتكون منها تلك الظاهرة، وذلك بتقسيم مساحة الدائرة إلى قطاعات يمثل كل قطاع منها جزءاً أو أحد مكونات الظاهرة . ويتم تحديد كل جزء من خلال ضرب الزاوية المركزية للدائرة والتي مقدارها 360° بحاصل قسمة الجزء المعني على مجموع الأجزاء كالآتي:

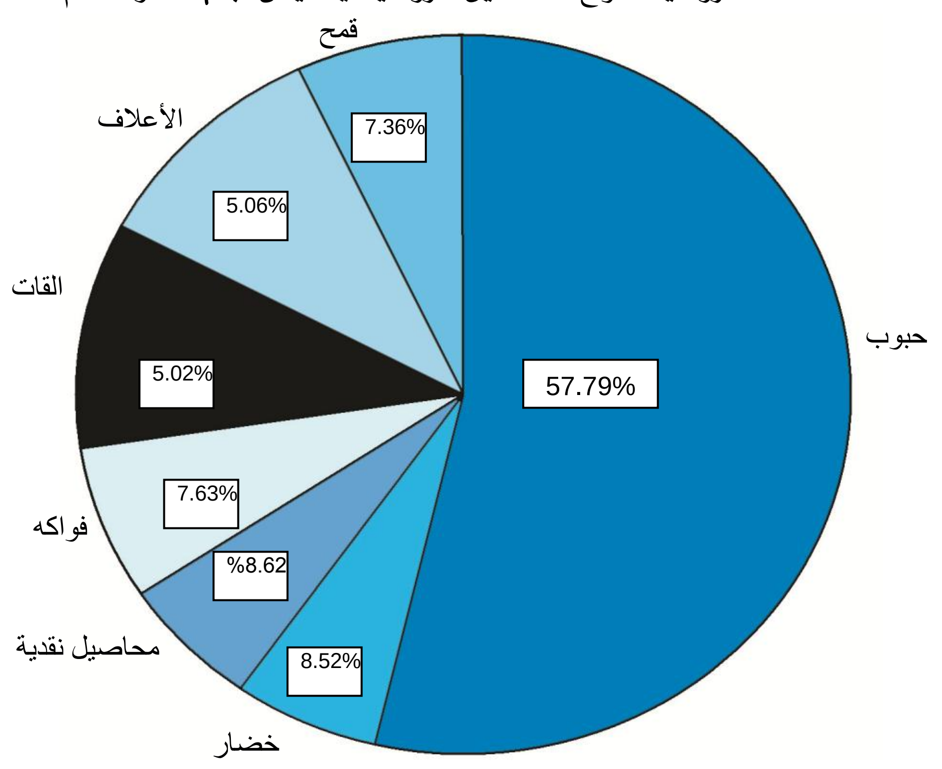
$$\text{مساحة الجزء المعني} = \frac{\text{قيمة الجزء}}{\text{مجموع قيم الأجزاء}} \times 360^{\circ}$$

والشـــكل(8.1) يمثـل المســاحات الزراعيــة في الــيمن (بالهكتــار) للعــام 2004م للمحاصيل الموضحة في الجدول التالي

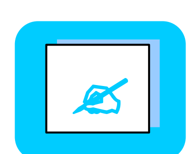
المحصول	قمح	الحبوب الأخرى	الخضار	المحاصيل النقدية	الفواكه	القات	الأعلاف
لمساحة بالهكتار	10.6743	63.5581	72.364	71.585	80.835	122.844	121.878

شكل (8.1)

المساحات الزراعية لأنواع المحاصيل الزراعية في اليمن (بالهكتار) للعام 2001م.



مثال(6.1):



أتضـح مـن أحـد البحـوث الاجتماعيـة أن هنـاك أسـرة تنفـق دخلـها الشهري البالغ 37000ريال شهرياً على النحو التالي: الغذاء 12000 ريـال، والسـكن 10000 ريـال، والملـبس 8000 ريـال، والـدواء 4000 ريال، وأخرى 3000 ريال.

المطلوب رسم دائرة بيانية لتمثيل هذه البيانات؟

الحل:

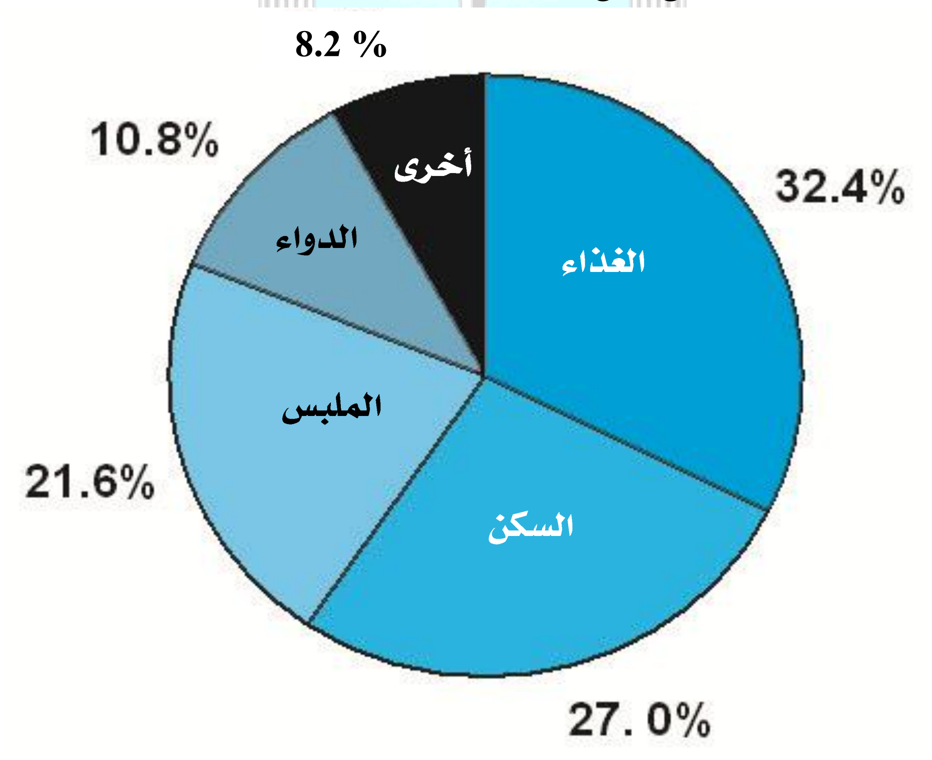
نقوم برسم الدائرة البيانية بعد حساب النسبة المئوية والزاوية المركزية لكل قطاع من هذه القطاعات حسب الجدول(3.1) والشكل رقم (9.1)

جدول (3.1)

البيان	المبلغ	النسبة المئوية	الزاوية المركزية
الغذاء	12000	32.4	116.76
السكن	10000	27.0	97.30
الملبس	8000	21.6	77.84
الدواء	4000	10.8	38.92
أخرى	3000	8.2	29.19
إجمالي الدخل	37000	100	360

شكل(9.1)

يوضح توزيع دخل الأسرة على البنود المذكورة



2.4 التوزيعات التكرارية (Frequency Distributions)

تناولنا في القسم السابق طرق العرض والتمثيل البياني للبيانات غير المبوبة، لبيانات محدودة، إلا أنه تواجهنا في الحياة العملية كميات كبيرة من البيانات قد لا نتمكن من عرضها بإحدى طرق عرض البيانات غير المبوبة، لأننا نحتاج إلى كتابة جداول تكرارية طويلة لنتمكن من عرض البيانات المطلوبة. وفي مثل هذه الحالة نقترح طريقة أخرى لعرض البيانات هي جداول التوزيع التكراري وهي إحدى الطرق التي نتمكن بواسطتها من تنظيم البيانات الكثيرة.

التوزيع التكراري: يعني تبويب عدد كبير من البيانات على شكل فئات تسمى بالفئات التكرارية، ثم تحديد عدد المشاهدات التي تقع ضمن كل من هذه الفئات ويطلق عليها بالتكرارات.

1.2.4 بناء التوزيع التكراري

أ. التبويب (Tabulation)

التبويب : هو وضع البيانات العددية في جداول وذلك بعد تقسيمها حسب صفاتها المشتركة.والغاية من هذه العملية هي اختصار البيانات إلى أصغر حيز يمكن أن يستوعبها وتمر عملية التبويب بثلاث مراحل:

1- تقسيم البيانات إلى مجموعات حسب صفاتها المشتركة، بحيث تنتمي كل مفردة إلى إحدى المجموعات فقط.

2- تحديد عدد مفردات كل مجموعة.

3- إنشاء الجدول الإحصائي ونقل البيانات إليه بحيث يظهر فيه عدد مفردات كل مجموعة

ويعرف الجدول الإحصائي بأنه: ترتيب البيانات في أسطر وأعمدة حسب صفاتها المشتركة بحيث نستطيع قراءة هذه البيانات بشكل عمودي أو أفقي أو كليهما معاً.

ونظراً لما يلعبه هذا النوع من التبويب من دور بالغ الأهمية لإعداد جدول تكراري يتم تصنيف البيانات إلى مجموعات.

ب- **الفئة (Class)** :

الفئة: هي **مجموعة جزئية محددة بدقة ووضوح وتحوي عدداً من القيم المتقاربة يعتقد الباحث أنها شبه متجانسة** ويفضل أن تكون هذه الفئات متساوية الطول ليكن من السهل إجراء المقارنات والقيام بالعمليات الحسابية.

ج- عدد الفئات:

لا توجد قاعدة ثابتة تحدد عدد الفئات المطلوبة لأن ذلك يتوقف على:

1- حجم البيانات.

2- الفرق بين أصغر وأكبر قيمة فيها.

3-مدى تجانسها.

4- مستوى الدقة المطلوبة.

ومن الاعتبارات التي يجب مراعاتها عند تحديد عدد الفئات:

1. يفضل أن يتراوح عدد الفئات بين 5 و 15.
2. ألا يكون عدد الفئات كبيراً فتضيع الفائدة من الجدول وهي اختصار البيانات.
3. ألا يكون عدد الفئات قليل جداً وبالتالي طول الفئة سيكون كبير فتضيع معالم التوزيع.

د. طول الفئة (Class Interval):

إن تحديد طول الفئة عملية تخضع لرأي الباحث حيث أن تحديد طول الفئة يعتمد على:

1- الفرق بين أكبر وأصغر قيمة في البيانات .

2- مدى تجانس البيانات .

3- عدد الفئات التي يرغب الباحث في إنشائها .

4-مستوى الدقة التي يرغب فيها الباحث.

حيث تستطيع تحديد طول الفئة من خلال قسمة المدى المطلق للبيانات على عدد الفئات ويفضل تقريب الكسر إلى أقرب عدد صحيح. أي أن:

$$\text{طول الفئة} = \frac{\text{المدى}}{\text{عدد الفئات}}$$

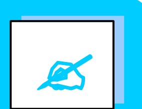
ملاحظة: مدى الفئة يساوي الحد الأعلى للفئة مطروحاً منه الحد الأدنى للفئة بصرف النظر عن نوع الظاهرة المدروسة.

هـ. حدود الفئات (Class Limits) :

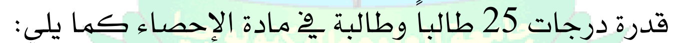
بعد تحديد طول الفئة يجب تحديد حدود كل فئة من الفئات بوضوح تام لضمان عدم التداخل بين الفئات. وهذا يتطلب التمييز بين نوعين من البيانات. بيانات عن الظاهرة المتصلة وبيانات عن الظاهرة المنفصلة ، لأن اختيار حدود الفئات يؤثر كثيراً على نتائج التحليل الإحصائي للبيانات والجدول الصحيح هو الجدول الذي نتائجه قريبة جداً من النتائج التي يمكن الحصول عليها من البيانات الخام قبل تبويبها.

وقد جرى العرف على قراءة حدود الفئات كما يلي: إذا تكررت كتابة الحد الأعلى للفئة كحد أدنى للفئة التالية فهذا يدل على أن مجال الفئة مفتوح من الأعلى أي أن المفردات التي قيمها تساوي الحد الأعلى للفئة لا تدخل ضمن الفئة ، أما إذا لم تتكرر كتابة الحد الأعلى للفئة كحد أدنى للفئة التالية فمجال الفئة مغلق من الطرفين.والمفردات التي قيمها مساوية للحد الأعلى للفئة تنتمي لهذه الفئة ، ويلجأ بعض المؤلفين إلى حل وسط بين الطريقتين وذلك بعدم ذكر الحد الأعلى للفئة.

ولبيان كيف تتم عملياً الخطوات السابقة نورد المثال (7.1):



مثال (7.1):



قدرة درجات 25 طالباً وطالبة في مادة الإحصاء كما يلي:

	61	53	42	47	35	10	19	17
	21	53	32	44	73	85	73	44
81	49	25	17	43	35	17	47	89

صمم جدول تكراري لهذه الدرجات.

الحل

لتصميم جدول توزيع تكراري ذي فئات نتبع الخطوات التالية:

1. نحدد اكبر قيمة وهي 89 وكذلك نحدد أصغر قيمة وهي 10 .
2. نحدد المدى وهو الفرق بين أكبر قيمة وأصغر قيمة وفي مثالنا نجد المدى 79.

$$89-10 = 79$$

3. اقترح عدد الفئات وليكن 8 (مع ملاحظة أنك أنت الذي قررت عدد الفئات ليكون 8 وبإمكانك اقتراح عدد آخر مثل 10 أو 7).

4. احسب طول الفئة من العلاقة:

$$\text{طول الفئة} = \frac{79}{8} = 9.9$$

وبما أن البيانات أعداد صحيحة يجب أن يكون طول الفئة عدداً صحيحاً وذلك بتقريبه إلى الأعلى فإن طول الفئة هو 10

5. ولتعيين حدود الفئات يمكن أن نبدأ بأي رقم أقل من أصغر رقم في البيانات أو مساوٍ له. وسيكون الحد الأدنى للفئة الأولى هو أصغر عدد في هذه البيانات وهو الرقم 10 وستكون الفئات كما يلي:

من 10 إلى 19

من 20 إلى 29

من 30 إلى 39

من 40 إلى 49

من 50 إلى 59

من 60 إلى 69

من 70 إلى 79

من 80 إلى 89

من الجدول نلاحظ أن :

الفئة 10-19 تحتوى على الدرجات 10، 17، 17، 17، 19

والفئة 20-29 تحتوي على الدرجات 21، 25 وهكذا....

والجدول رقم (4.1) يبن عملية تفريغ الدرجات

UNIVERSITY OF SCIENCE & TECHNOLOGY
جامعة العلوم والتكنولوجيا

جدول (4.1)

التوزيع التكراري لدرجات الطلاب في مادة الإحصاء

الفئات	الفرز	التكرار(f_i)
10 - 19	~~////~~	5
20 – 29	//	2
30 – 39	////	3
40 – 49	// ~~////~~	7
50 – 59	//	2
60 – 69	/	1
70 – 79	//	2
80 – 89	///	3

وعلى الرغم من أن الفئات تضم كافة المعطيات عند تبويبها، إلا أننا نلاحظ أن هذه الفئات غير متصلة ببعضها، ولمعالجة هذه المسألة التي يكون لها تأثير مباشر على التوزيعات الاحتمالية، نحسب ما يعرف بالحدود الحقيقية للفئات من خلال العلاقة:

الحد الأدنى الحقيقي = الحد الأدنى الاسمي – 0.5

الحد الأعلى الحقيقي = الحد الأعلى الاسمي + 0.5

فتكون الحدود الحقيقية للفئات كما في الجدول (5.1)

و. مركز الفئة:

هو القيمة التي تتوسط بين الحد الأدنى والحد الأعلى للفئة.

وتتبع أهمية مركز الفئة من أن كل مفردة في أي فئة يفترض أنها مساوية لمركز تلك الفئة. لأن مركز الفئة هو قيمة واحدة تعبر عن الفئة بأسرها. ويمكن حساب مركز الفئة بإحدى الطرق التالية:

$$\text{مركز الفئة } (X_i) = \frac{\text{الحد الأدنى + الحد الأعلى}}{2}$$

$$\text{أو مركز الفئة } (X_i) = \text{قيمة الحد الأدنى للفئة} + \frac{1}{2}\ \text{مدى الفئة}$$

$$\text{أو مركز الفئة } (X_i) = \text{قيمة الحد الأعلى للفئة} - \frac{1}{2}\ \text{مدى الفئة}$$

فنجد أن مركز الفئة الأولى (10-19) هو $=10+(9)\times\frac{1}{2}=14.5$

جدول(5.1)

مركز الفئة	الحدود الحقيقية للفئات	الحد الأعلى الحقيقي	الحد الأدنى الحقيقي	حدود الفئات
14.5	9.5 – 19.5	19.5	9.5	10-19
24.5	19.5 – 29.5	29.5	19.5	20-29
34.5	29.5 – 39.5	39.5	29.5	30-39
44.5	39.5 – 49.5	49.5	39.5	40-49
54.5	49.5 – 59.5	59.5	49.5	50- 59
64.5	59.5 – 69.5	69.5	59.5	60- 69
74.5	69.5 – 79.5	79.5	69.5	70- 79
84.5	79.5 – 89.5	89.5	79.5	80-89

ويفضل عند بناء التوزيع التكراري مراعاة النقاط الآتية:

1. أن تكون الفئات منفصلة عن بعضها (غير متداخلة)
2. أن تكون الفئات متساوية الطول قدر الإمكان.
3. أن تكون الفئات كافية لاحتواء جميع البيانات.

ولا يفوتنا في هذا الصدد أن نشير إلى أن الجداول التكرارية تتنوع حسب الغاية من إعدادها وكذلك حسب عدد المتغيرات المعروضة في الجدول ومنها:

1- **الجدول المنتظم و الجدول غير المنتظم**: يكون التوزيع منتظماً إذا كانت أطوال فئاته متساوية كما في الجدول السابق. ويكون جدول التوزيع غير منتظم إذا كانت أطوال فئاته غير متساوية.

2- **الجدول المغلق والجدول المفتوح**: يدعى جدول التوزيع التكراري مغلقاً إذا علمت حدود كل الفئات، ويكون مفتوحاً إذا كانت بداية الفئة الأولى أو نهاية الفئة الأخيرة أو كليهما معاً غير محددة ومن الأفضل دوماً قفل الجدول من طرفيه لتسهيل العمليات الحسابية.

2.2.4 جداول التوزيع التكرارية المتجمعة:

comulatine frequency distribution

في حالات عديدة ينصب الاهتمام ليس على عدد التكرارات في فئة معينة، ولكن على العدد الذي يزيد أو يقل عن قيمة معينة فقد ترغب في معرفة عدد المفردات التي تكون قيمتها أقل أو أكبر من قيمة معينة ففي المثال السابق قد يهمنا مثلاً معرفة عدد الطلاب الذين درجاتهم أقل من 50 درجة فنجد أنهم 17 طالب وهذا العدد هو مجموع التكرارات بالفئات الأربع الأولى بالجدول التكراري ومجاميع هذه التكرارات تسمى بالتكرارات المتجمعة وهناك نوعين من التكرارات المتجمعة هي :

1- **تكرار متجمع صاعد**: وهي التكرارات التي يبدأ تجميعها من الأعلى باتجاه الأسفل حيث يتم البدء بأصغر الفئات ووضع التكرارات أمام الفئات بحيث يتضمن التكرار المقابل لكل فئة مجموع التكرارات للفئات الأقل منها. ويلاحظ على الجدول التكراري المتجمع الصاعد أنظر (جدول6.1) ما يلي:

1- أضفنا فئة في بداية الجدول تكرارها صفر لغرض الرسم البياني.

2- نبدأ فيه بتجميع التكرارات من جهة الفئات ذات الحدود الصغيرة .

3- التكرار المتجمع الصاعد المناظر للفئة الأولى يساوي تكرار الفئة الأولى.

4- التكرار المتجمع الصاعد المناظر للفئة الأخيرة يساوي مجموع التكرارات.

2- **تكرار متجمع نازل أو هابط**: وهي التكرارات التي يبدأ تجميعها من الأسفل إلى الأعلى حيث يتم البدء بأكبر الفئات ووضع التكرارات أمام الفئات بحيث يتضمن التكرار المقابل لكل فئة مجموع التكرارات للفئات الأكبر منها. ويلاحظ على الجدول التكراري المتجمع الهابط ما يلي:

1- أضفنا فئة في نهاية الجدول تكرارها صفر لغرض الرسم البياني

2- نبدأ فيه بتجميع التكرارات من جهة الفئات ذات الحدود الكبيرة..

3- التكرار المتجمع الهابط المناظر للفئة الأولى يساوي مجموع التكرارات.

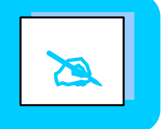
4- التكرار المتجمع الهابط المناظر للفئة الأخيرة يساوي تكرار الفئة الأخيرة.

ويمكن حساب التكرار المتجمع النسبي أمام كل فئة ممثلة بالنهاية العليا وذلك بقسمة التكرار المتجمع للفئة المعينة على مجموع لتكرار.

مثال(8.1):

كون جدول التوزيع التكراري الصاعد والهابط لبيانات المثال (7.1)

إذا علم أن :



فئات	10-	20-	30-	40-	50-	60-	70-	80 - 89
تكرارات	5	2	3	7	2	1	2	3

الحل :

الجدولين (6.1)، (7.1) يوضحان التوزيعين التكراريين الصاعد والنازل لبيانات المثال (3.2.4)

جدول (6.1) تكرار متجمع صاعد

الحدود العليا للفئات	تكرار متجمع صاعد
أقل من 10	0
~ ~ 20	5
~ ~ 30	7
~ ~ 40	10
~ ~ 50	17
~ ~ 60	19
~ ~ 70	20
~ ~ 80	22
~ ~ 90	25

جدول (7.1) تكرار متجمع هابط

الحدود الدنيا للفئات	تكرار متجمع هابط
أكبر من 9	25
~ ~ 19	20
~ ~ 29	18
~ ~ 39	15
~ ~ 49	8
~ ~ 59	6
~ ~ 69	5
~ ~ 79	3
~ ~ 89	0

3.2.4 التوزيع التكراري النسبي: Relative Frequencies

بعد بناء التوزيع ومعرفة تكرار كل فئة فيه قد يكون الاهتمام منصباً على نسبة الأفراد في كل فئة لا على العدد بحد ذاته. فعلى سبيل المثال : إذا قرأت في جدول توزيع تكراري يمثل أعداد الطلاب الملتحقين في جامعة العلوم والتكنولوجيا لعام 2007/2006م فقد تتساءل ما نسبة الطلبة الملتحقين في كلية الطب البشري وفي هذه الحالة فإن اهتمامك لا يكون منصباً على عدد الطلبة الملتحقين في الطب البشري بل على نسبة الطلبة الملتحقين في الطب البشري من مجموع الملتحقين

بالجامعة لنفس العام. وهذا يقودك إلى تعريف التكرار النسبي. والذي يعرف بأنه **عبارة عن نسبة ما يشكله تكرار كل فئة إلى مجموع التكرارات**. وتحسب النسب من خلال قسمة تكرار كل فئة على مجموع التكرارات ثم نضربها في 100 ويحسب من العلاقة.

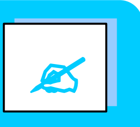
$$\text{التكرار النسبي} = \frac{\text{تكرار الفئة}}{\text{مجموع التكرارات}}$$

أما إذا ضربنا التكرار النسبي بـ 100 فإننا نحصل على التكرار السنوي حيث أن مجموع التكرار المئوي=100

جدول(8.1) التكرار النسبي لبيانات المثال (7.1)

الفئات	f_i	التكرار النسبي	التكرار المئوي ٪
10-19	5	0.20	20
20-29	2	0.08	8
30-39	3	0.12	12
40-49	7	0.28	28
50- 59	2	0.08	8
60- 69	1	0.04	4
70- 79	2	0.08	8
80-89	3	0.12	12
المجموع	25	1	100

مثال(9.1):



بفرض أن لدينا البيانات التالية عن أطوال خمسين طالب من طلاب كلية الشرطة والمطلوب هو تصميم جدول توزيع تكراري لهذه الأطوال.

176	155	179	158	165	190	165	171	177	171
171	165	183	157	185	165	171	171	173	171
178	150	159	161	187	171	163	170	177	185
177	171	183	165	190	177	164	169	175	171
182	160	171	181	167	183	166	172	171	177

الحل

المدى = أكبر قيمة – أقل قيمة

$$190 - 150 = 40$$

نقترح عدد الفئات يساوي سبع فئات فيكون

طول الفئة $\frac{40}{7} = 5.7$

نقرب طول الفئة إلى العدد 6

نكتب حدود الفئات مبتدئين بأصغر عدد في البيانات كحد أدنى للفئة الأولى ونضيف عليه طول الفئة على التوالي فنحصل على حدود جميع الفئات كما هي موضحة في الجدول رقم (9.1)

جدول(9.1)

يوضح حدود الفئات ومراكز ها والتكرار النسبي لبيانات المثال (9.1)

التكرار النسبي ٪	مراكز الفئات	التكرار f_i	الفئات
4	153	2	150 - 156
10	159	5	156 – 162
18	165	9	162 – 168
30	171	15	168 – 174
18	177	9	174 – 180
14	183	7	180 – 186
6	189	3	186 – 192
٪100.00	--	50	المجموع

تدريب (6)



1- عرف كلاً من: الفئة ، مركز الفئة ، التبويب، الجدول الإحصائي.

2- ما هي الاعتبارات التي يجب مراعاتها عند تحديد عدد الفئات ؟

3- ما أنواع الجداول التكرارية ؟

تدريب (7)



1- ما نوع الظاهرة :

من 130إلى 135 ، من 136إلى 141

ثم أوجد 1- طول الفئة 2- مركز الفئة 3- مدى الفئة

2- الإعداد التالية تمثل أفراد 30 أسرة ضع هذه البيانات في صورة جدول تكراري متجمع صاعد (طول الفئة 2).

5	6	4	3	3	6	2	3	4	5
8	6	7	2	6	2	3	4	7	8
4	6	3	6	4	5	6	6	5	7

أسئلة التقويم الذاتي (4) :

أ- أكمل الفراغات التالية بالعبارات أو الكلمات المناسبة:

1- يمكن عرض البيانات الإحصائية بعدة طرق منها ....و....و...

2- تمر عملية التبويب بثلاث مراحل هي ......و......و...........

3- المدى هو الفرق بين ....و.........................

4- من طرق التعبير عن الفئات ....و...............

5- تستخدم الأعمدة البيانية الأحادية لعرض بيانات تخص ظاهرة.........
بينما تستخدم الأعمدة المركبة لعرض ....لظاهرة واحدة

ب-بين ما هو أسلوب العرض البياني الممكن استخدامه لعرض المتغيرات التالية:

- عدد المستشفيات حسب محافظات الجمهورية لعام 2005
- الدخل الشهري لـ500 عامل
- عدد الأشجار المثمرة حسب أنواعها في محافظة صعدة عام 2005

ج-حدد نوع الظاهرة في :

من 130 إلى 136 ، من 136 إلى 142

ثم أوجد :

1) طول الفئة . 2) مركز الفئة.

?

3.4 التمثيل البياني للتوزيعات التكرارية:

(Graphical presentation of Frequency Distributions)

التمثيل البياني هو تخطيط يعرض المعلومات بشكل مرئي مما يساعد في فهم الأرقام والمقارنة بينها . ويتم عرض التوزيعات التكرارية بيانياً غالباً بإحدى الطرق الآتية:

1- المدرج التكراري .

2- المضلع التكراري .

3- المنحنى التكراري .

1.3.4 المدرج التكراري Histogram

يتكون من مجموعة من المستطيلات المتلاصقة على المحور الأفقي قاعدة كل منها تساوي طول الفئة وارتفاعها يساوي تكرار هذه الفئة. فهو يشبه الأعمدة البيانية، إلا أنه يختلف عنها في كون مستطيلاته متلاصقة، ولرسم المدرج التكراري نتبع الخطوات التالية:

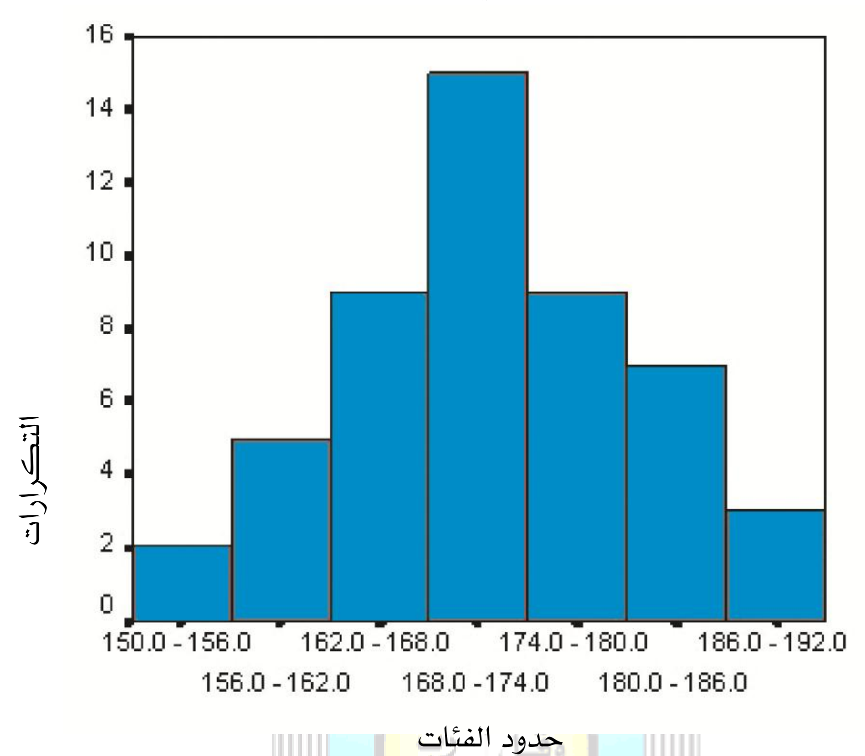
1- ارسم محورين متعامدين بحيث يخصص المحور الأفقي للفئات والمحور الرأسي للتكرارات.

2- نقسم المحور الأفقي إلى أقسام متساوية كل قسم منها عبارة عن طول الفئة.

3- ارسم مستطيلاً على كل فئة قاعدته تساوي طول الفئة وارتفاعه يساوي تكرار هذه الفئة. كما هو موضح في الشكل (10.1) والذي يمثل المدرج التكراري للتوزيع المعطى في جدول(9.1) . أي أن مجموع مساحات المستطيلات يكون مساوياً للمجموع الكلي للتكرارات ويستخدم بشكل رئيس في عرض البيانات الكمية.

شكل (10.1)

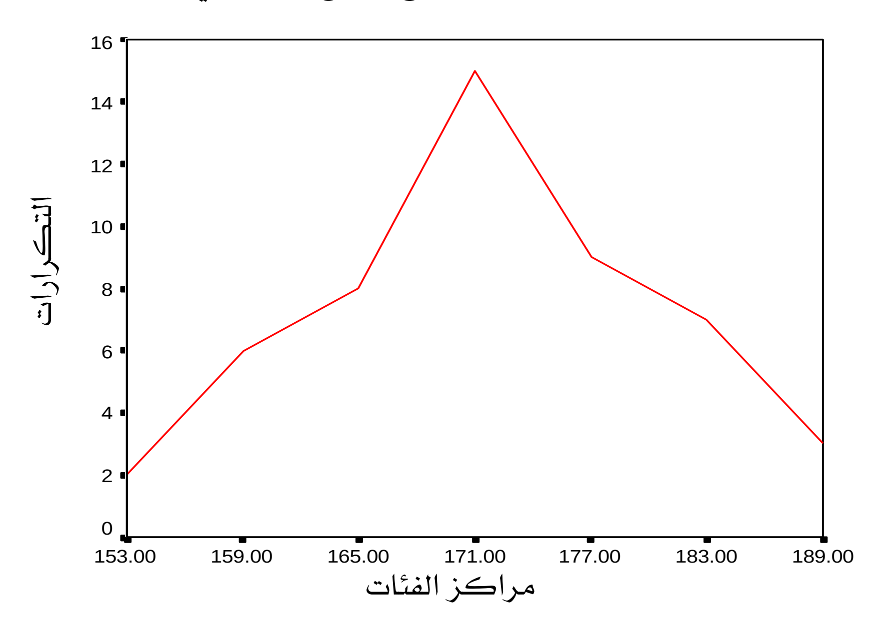
يوضح المدرج التكراري لبيانات الجدول (9.1)



2.3.4 المضلع التكراري Frequency Polygon

هو عبارة عن خط منكسر يصل بين النقاط التي إحداثياتها هي مراكز الفئات أفقياً (على المحور الأفقي) والتكرار رأسيا(على المحور العمودي)، ومن ثم التوصيل بين نهايات النقاط للأعمدة التكرارية بخطوط مستقيمة، بعد إضافة فئتين عند بداية ونهاية المحور الأفقي وبتكرار مقداره صفر كما يتضح من الشكل(11.1) والذي يوضح المضلع التكراري للتوزيع التكراري في الجدول (9.1)

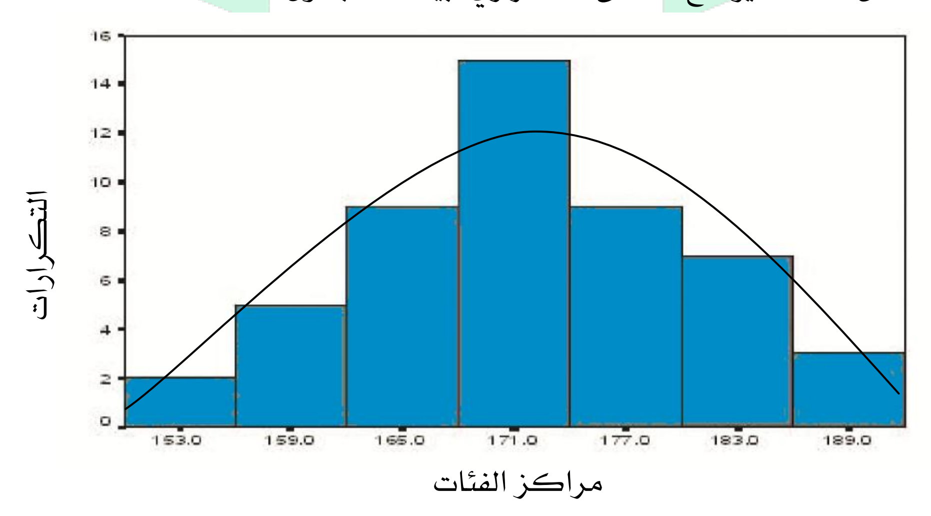
شكل(11.1) يوضح المضلع التكراري



3.3.4 المنحنى التكراري Frequency Curve

وبتمهيد المضلع محصلة منحنى بدلاً من الخطوط المتكسرة نحصل على ما يسمى بالمنحنى التكراري. ويتم رسم المنحنى التكراري بنفس طريقة رسم المضلع التكراري حيث توصل النقاط بخط ممهد مستمر(غير متكسر) يمر بأغلب النقاط ويتوسط بقيتها خير توسط كما هو موضح بالشكل (12.1) والذي يمثل بيانات الجدول(9.1).

شكل (12.1) يوضح المنحنى التكراري لبيانات الجدول (9.1)

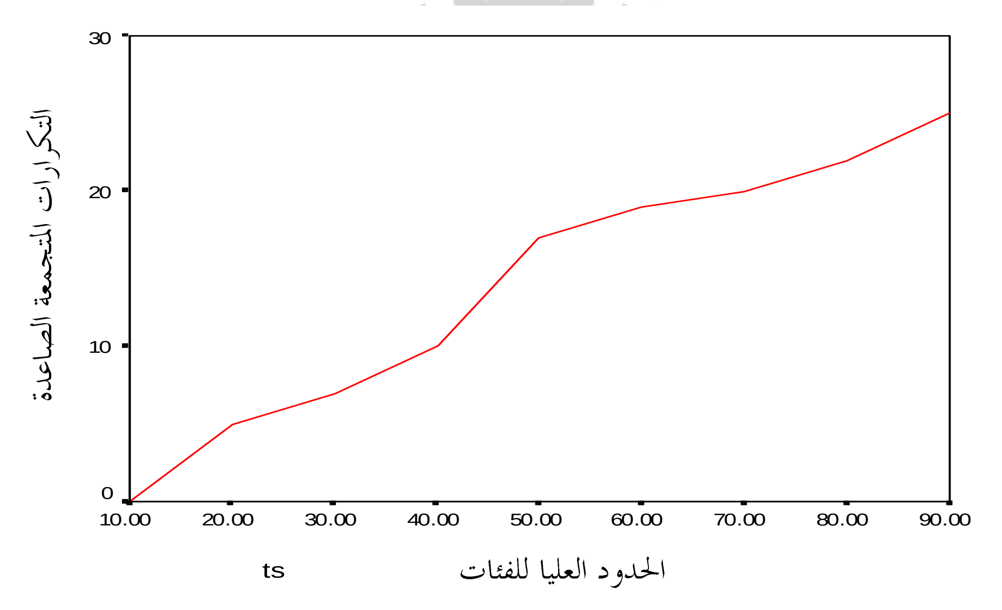


4.3.4 تمثيل التوزيعات التكرارية المتجمعة بيانياً: Graphic presentations

بينا فيما سبق كيفية تكوين الجداول التكرارية الصاعدة والهابطة من الجدول التكراري البسيط. ولعرض بيانات الجداول التكرارية المتجمعة بيانياً نرسم محورين متعامدين ونخصص المحور الأفقي لحدود الفئات والمحور الرأسي للتكرارات المتجمعة. ولتمثيل بيانات الجدول التكراري المتجمع الصاعد جدول(4.1) بيانياً نخصص المحور الأفقي للحدود العليا للفئات والمحور الرأسي للتكرار المتجمع الصاعد.والشكل (13.1) يمثل المنحنى التكراري المتجمع الصاعد لبيانات الجدول (4.1).

شكل(13.1)

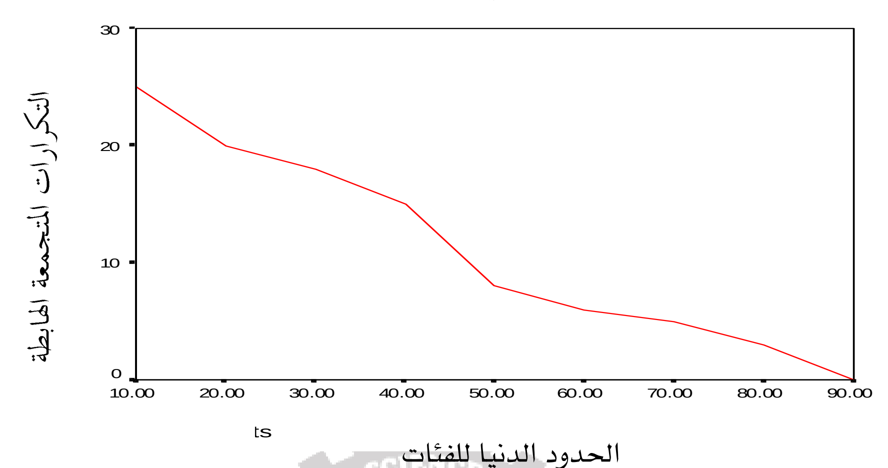
المنحنى التكراري المتجمع الصاعد لبيانات الجدول (4.1).



ولتمثيل بيانات الجدول التكراري المتجمع الهابط جدول(4.1) بيانياً نخصص المحور الأفقي للحدود الدنيا للفئات والمحور الرأسي للتكرار المتجمع الهابط. والشكل (14.1) يمثل المنحنى التكراري المتجمع الهابط لبيانات الجدول (4.1).

الشكل(14.1)

المنحنى التكراري المتجمع الهابط للجدول(4.1)



ويمكن رسم المنحنيين المتجمعين الصاعد والهابط على شكل واحد وذلك بتخصيص المحور الأفقي لحدود الفئات –العليا والدنيا- والمحور الرأسي للتكرار المتجمع الصاعد والهابط (أنظر الوحدة الثانية عند حساب الوسيط بيانياً)

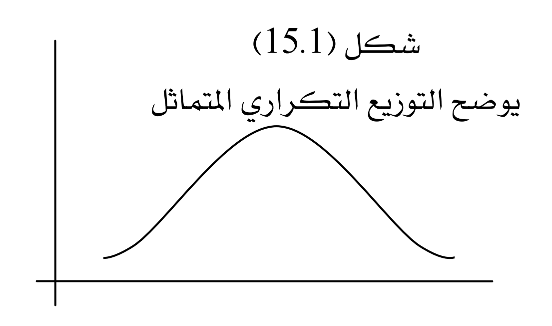
4.4 التماثل والالتواء والتفرطح في التوزيعات التكرارية:

من أهم الخصائص أو الصفات التي تميز التوزيعات صفة التماثل والالتواء والتفرطح، ويجدر بنا ونحن ندرس المنحنى التكراري أن نستعرض هنا التوزيعات المتماثلة والتوزيعات غير المتماثلة. فما هو التوزيع المتماثل ؟

يقال للتوزيع متماثل عندما يتطابق نصفيه عند محور عمودي بحيث ينطبق أحد نصفيه على الآخر تمام الانطباق وتسمى النقطة التي تقع على المحور الأفقي والتي يتم الطي عندها (نقطة التماثل)، كما يسمى الإحداثي العمودي المار بها (محور التماثل)للمنحنى والشكل (15.1) يوضح التوزيع التكراري المتماثل والذي تكون متوسطاته متساوية .

شكل (15.1)

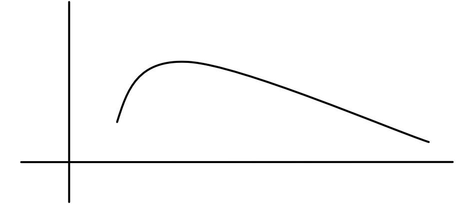
يوضح التوزيع التكراري المتماثل



وعندما لا يتطابق جانبي التوزيع يقال عنه توزيع ملتوي فعندما يكون الالتواء باتجاه اليمين يقال عنه توزيع موجب الالتواء شكل (16.1) حيث يصبح طرفه الأيمن أطول من طرفه الأيسر .

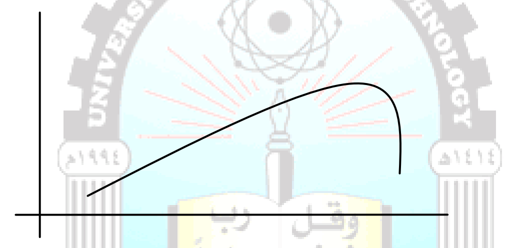
شكل (16.1)

يوضح التوزيع التكراري الموجب الالتواء



شكل (17.1)

يوضح التوزيع التكراري السالب الالتواء



وهناك أنواع أخرى من الرسوم البيانية منها الخرائط البيانية التي توضح الكثافة السكانية وتركزها أو الثروة المعدنية أو الحيوانية وشكل الساق والورقة...الخ.





تدريب (8)

ما الفرق بين المضلع التكراري والمدرج التكراري

أسئلة التقويم الذاتي (5) :

1) عدد طرق تمثيل التوزيعات التكرارية بيانياً.

2) حدد طبيعة التوزيع للبيانات التالية من خلال رسم المنحنى التكراري.

عدد الأسر f_i	فئات الدخل الشهري بآلاف الريالات
10	5 – 9
20	9 – 13
30	13 – 17
40	17 – 25
50	25 – 37
150	المجموع

ثم أوجد التكرار المتجمع الصاعد والهابط لبيانات الجدول السابق.



5- الخلاصة:

ركزت هذه الوحدة على المفاهيم الأساسية لعلم الإحصاء من حيث تعريفه وأنواعه ودوره في الارتقاء بالعملية الإنتاجية والإدارية. حيث أوضحت الوحدة أن علم الإحصاء ينقسم إلى قسمين هما :الإحصاء الوصفي والإحصاء الاستدلالي، وأن المتغيرات الإحصائية تنقسم إلى متغيرات كمية وهي المتغيرت التي تأخذ قيماً عددية، ومتغيرات نوعية وهي التي تصنف المشاهدات في مجموعات كل مجموعة تشترك في خاصية محددة.

وقد بينا أن نوع المقياس الإحصائي يعتمد على نوع المتغير المدروس فالمتغيرات النوعية تقاس بمقياس ترتيبي، ومقياس أسمي بينما لمتغيرات الكمية تقاس بمقياس فئوي، ومقياس نسبي.وأن طرق جمع البيانات الإحصائية تعددت بسبب تعدد طبيعة المجتمعات الإحصائية واختلاف البيانات ومنها طريقة المشاهدة أو الملاحظة، والاستبانة، والمقابلة الشخصية، والهاتف.

وبينا أيضاً أن المجتمع الإحصائي: يعني جميع المفردات التي تتمتع بصفة ما أو خاصية ما. وأن المجتمع الإحصائي يمكن تحديده زماناً ومكاناً، وأن مفهوم المجتمع الإحصائي هو مفهوم مرن حيث يمكن أن يوسع أو العكس بتوسع الزمان والمكان أو كليهما.كما أوضحنا أن أساليب جمع البيانات تنقسم إلى نوعين وهما :

أسلوب المسح الشامل: الذي يعتمد على جميع مفردات المجتمع الإحصائي دون استثناء ويمتاز بدقة النتائج، إلا أن له بعض العيوب ومنها أنه يحتاج إلى وقت طويل وجهد كبير وتكاليف عالية.

أسلوب المسح بالعينة والذي يعتمد على مجموعة جزئية من المجتمع الإحصائي .

وأن أسلوب المسح بالعينة يفضل على أسلوب المسح الشامل لعدة أسباب أهمها: قلة التكاليف، اتساع نطاق الدراسة، زيادة سرعة إنجاز الدراسة، الطبيعة الجغرافية وأن المجتمع الإحصائي قد يكون غير قابل للعد.

.وقد اهتمت هذه الوحدة بدراسة وشرح العينات العشوائية وطرق اختيارها وكذا طرق المعاينة العشوائية والتي منها المعاينة العشوائية البسيطة، والمنتظمة ،والطبقية، والعنقودية.

كما تم استعراض الأساليب المختلفة لعرض البيانات جدولياً وبيانياً والتي من أهمها: التوزيعات التكرارية البسيطة، وأنواع الجداول التكرارية. وقد حاولنا في هذه الوحدة تسليط الضوء على مفهوم العرض البياني **والذي يمثل تخطيط يعرض المعلومات بشكل مرئي مما يساعد على فهم الأرقام والمقارنة بينها**، ويتم عادة اختيار طريقة التمثيل البياني بحيث تناسب المقياس الإحصائي الذي قيست به البيانات المطلوب تمثيلها بيانياً.

كما استعرضت الوحدة مفهوم الأعمدة البيانية والتي تنقسم إلى الأعمدة البيانية البسيطة الأحادية والمتلاصقة والمركبة بهدف ملاحظة تطور الظاهرة خلال تعاقب الزمن. أو باستخدام الرسوم الدائرية بهدف إبراز الأجزاء التي تتكون منها تلك الظاهرة.



6- لمحة مسبقة عن الوحدة الدراسية الثانية:

عزيزي الدارس سبق أن ذكرنا أن العرض البياني لبيانات التوزيع التكراري يمدنا بصورة مرئية تساعد على فهمها ، إلا أنه على الرغم من أهمية العرض البياني فإنه لا يستجيب للأساليب الرياضية اللازمة لتحليل البيانات إحصائيا . لذا فإن الوحدة الدراسية الثانية سوف تتناول بعض المقاييس الإحصائية المناسبة لتحليل البيانات ومنها مقاييس النزعة المركزية، ومقاييس التشتت ، والتماثل والالتواء لتوضيح طبيعة التوزيع .



7- إجابات التدريبات:

تدريب(1):

الإحصاء الوصفي: يختص بطرق جمع البيانات ووصفها لتكون بصيغة مفهومة وذات مدلول دون الوصول إلى استنتاجات واستدلالات خاصة. أي هو التعامل مع البيانات الإحصائية دون التعميم. فهو عادة خطوة تسبق الإحصاء الاستدلالي.

الإحصاء الاستدلالي: يتعلق الإحصاء الاستدلالي بطرق تحليل وتفسير وتقدير واستخلاص الاستنتاجات، بالاعتماد على جزء (عينة) من المجتمع الأصلي للتوصل إلى قرارات تخص المجتمع الإحصائي. فهو الذي يتعامل مع التعميم والتقدير والتنبؤ.

تدريب(2):

المتغير	نوعه	نوع المقياس المناسب
مستوى الدخل	نوعي	ترتيبي
الوزن	كمي	نسبي
نوع العمل	نوعي	أسمي
حوادث الطرق	كمي	فئوي

تدريب(3):

المجتمع المحدد : يكون المجتمع محدد عندما نستطيع حصر عدد مفرداته وبالتالي نستطيع جمع بيانات عنها مثل سكان مدينة ذمار أو عدد المدخنين في مدينة ما و عدد الأطفال دون سن الخامسة في بلد ما.

الحصر الشامل (Survey): يقصد به جمع المعلومات عن الظاهرة موضوع الدراسة من جميع مفردات الظاهرة المدروسة دون استثناء مثل التعداد العام للمساكن والسكان.

أسلوب العينة: العينة يقصد بها جزء من مجتمع تؤخذ منه بطريقة أو بأخرى بحيث تكون العينة ممثلة للمجتمع تمثيلاً دقيقاً، وتدرس بهدف التعرف على خصائص وصفات المجتمع المأخوذة منه. أي أن استخدام العينة هو تعميم من الجزء إلى الكل.

المعاينة الإحصائية: هي استخدام الطرق والنظريات الإحصائية العلمية في أخذ العينة ودراستها. مثل مرضى السرطان في مدينة ما.

تدريب(4):

المجتمع الإحصائي: هو جميع مفردات المجتمع تحت الدراسة مثل سكان اليمن، وأعضاء هيئة التدريس في كلية العلوم الإدارية، وعدد الموظفين في الجامعة.

إطار المعاينة: هو قائمة تحتوي أو تتضمن مفردات المجتمع التي تشترك في صفة ما مثل سكان اليمن الذين لديهم هاتف ثابت.

تدريب(5):

$$\frac{n}{N}=\frac{n_1}{N_1}=\frac{n_2}{N_2}=\frac{n_3}{N_3}$$

$$\frac{120}{16000}=\frac{n_1}{9000}=\frac{n_2}{6000}=\frac{n_3}{1000}$$

$n_1 = 68$

$n_2 = 45$

$n_3 = 7$

تدريب(6):

الفئة: هي مجموعة جزئية محددة بدقة ووضوح وتحوي عددا من القيم المتقاربة يعتقد الباحث أنها شبه متجانسة.

التبويب: هو وضع البيانات العددية في جداول وذلك بعد تقسيمها حسب صفاتها المشتركة بهدف اختصار البيانات إلى أصغر حيز ممكن أن يستوعبها.

الجدول الإحصائي: يعرف الجدول الإحصائي بأنه ترتيب البيانات في أسطر وأعمدة حسب صفاتها المشتركة بحيث نستطيع قراءة هذه البيانات بشكل رأسي أو أفقي أو كليهما.

مركز الفئة: هو القيمة التي تتوسط بين الحد الأدنى والحد الأعلى للفئة.

تدريب(7):

نوع الظاهرة منفصلة .

طول الفئة = (الحد الأعلى للفئة – الحد الأدنى للفئة) + 1

= (135 - 130) + 1 = 5+1= 6 لأن الظاهرة منفصلة

مركز الفئة هو 132.5

مدى الفئة = الحد الأعلى للفئة – الحد الأدنى للفئة = 135 - 130 = 5

جامعة العلوم والتكنولوجيا UNIVERSITY OF SCIENCE & TECHNOLOGY

2-جدول التكرار التجمع الصاعد

الفئات	التكرار	التكرار المتجمع الصاعد
2-4	8	8
4-6	9	17
6-8	13	30
	30	

تدريب (8)

المدرج التكراري: هو عبارة عن تمثيل تكرار كل فئة من فئات التوزيع التكراري ذي الفئات المتساوية بمستطيل حدود قاعدته هي الحدود الحقيقية لتلك الفئة، يتناسب ارتفاعه مع تكرارها.

المضلع التكراري: هو عبارة عن خط منكسر يصل بين النقاط التي إحداثياتها هي مراكز الفئات أفقياً والتكرار رأسياً والتوصيل بين نهايات النقاط بخطوط مستقيمة.



8- قائمة المصطلحات:

- **المتغيرات (Variables):** هو خاصية قابلة للتغير من مشاهدة إلى أخرى في المجتمع الإحصائي. فكل ظاهرة تأخذ قيماً عددية تسمى متغيراً.

- **المجتمع الإحصائي (Statistical Population):** هو جميع المفردات التي تتمتع بصفة ما أو خاصية مشتركة وهذه المفردات قد تكون بشراً أو أشياء أخرى.

- **العينات غير الاحتمالية (Non Probability Sampling):** يقصد بالعينات غير الاحتمالية تلك التي يتم اختيارها وفق معايير يحددها الباحث ويعتقد أنها ستؤدي إلى الحصول على عينة ممثلة للمجتمع.

- **العينات الاحتمالية (Probability Sampling):** هي عينات يتم اختيارها بأسلوب يسمح بتحديد احتمال ظهور كل مفردة من مفردات المجتمع في العينة ويكون ذلك الاحتمال لا يساوي الصفر.

- **طرق المعاينة (Sampling Methods):** هي الطرق التي يتم بواسطتها اختيار العينة

- **التمثيل البياني:** هو تخطيط يعرض المعلومات بشكل مرئي مما يساعد في فهم الأرقام والمقارنة بينها.

- **المدرج التكراري (Histogram):** يتكون من مجموعة من المستطيلات المتلاصقة على المحور الأفقي قاعدة كل منها تساوي طول الفئة وارتفاعها يساوي تكرار هذه الفئة. فهو يشبه الأعمدة البيانية ، إلا أنه يختلف عنها في كون مستطيلاته متصلة

- **وحدة المعاينة أو المشاهدة (Unit of Observation):** هي الوحدة في المجتمع التي نقوم بمشاهدة إحدى خواصها.

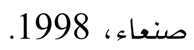
9- الملاحق

ملحق (1) جزء من الأرقام العشوائية

93108	77033	68325	10160	38667	62441	87023	94372	06164	30700
28271	08589	83279	48838	60935	70541	53814	95588	05832	80235
21841	35545	11148	34775	17308	88034	97765	35959	52843	44895
22025	79554	19698	25255	50283	94037	57463	92925	12042	91414
09210	20779	02994	02258	86978	85092	54052	18354	20914	28460
90552	71129	03621	20517	16908	06668	29916	51537	93658	29525
01130	06995	20258	10351	99248	51660	38861	49668	74742	47181
22604	56719	21784	68788	38358	59827	19270	99287	81193	43366
06690	01800	34272	65497	94891	14537	91358	21587	95765	72605
59809	69982	71809	64984	48709	43991	24987	69246	86400	29559
56475	02726	58511	95405	70293	84971	06676	44075	32338	31980
02730	34870	83209	03138	07715	31557	55242	61308	26507	06186
74482	33990	13509	92588	10462	76546	46097	01825	20153	36271
19793	22487	94238	81054	95488	23617	15539	94335	73822	93481
19020	27856	60526	24144	98021	60564	46373	86928	52135	74919
69565	60635	65709	77887	42766	86698	14004	94577	27936	47220
69274	23208	61035	84263	15034	28717	76146	22021	23779	98562
83658	14204	09445	41081	49630	34215	89806	40930	97194	21747
78612	51102	66826	40430	54072	62164	68977	95583	11765	81072
14980	74158	78216	38985	60838	82836	42777	85321	90463	11813
63172	25010	29405	91554	75195	51183	65805	87525	35952	83204
71167	37984	52737	06869	38122	95322	41356	19391	96787	64410
78530	56410	19195	34434	83712	50397	80920	15464	81350	18673
98324	03774	07573	67864	06497	20758	83454	22756	83959	96347
55793	30055	08373	32652	02654	75980	02095	87545	88815	80086
05674	34471	61967	91266	38814	44728	32455	17057	08339	93997
15650	22242	07592	22078	73628	60902	41561	54608	41023	98345
66750	19609	70358	03622	64898	82220	69304	46235	97332	64539
42320	74314	50222	82339	51564	42885	50482	98501	02245	88990
73725	73818	15470	04914	24936	65514	56633	72030	30856	85183
97546	02188	46373	21486	28211	08155	23486	66134	88799	49496
23569	52162	38444	42004	78011	16909	94194	79732	47114	23919
36048	93973	82596	28739	86985	58144	65007	08786	14826	04896
40455	36702	38965	56042	80023	28169	04174	65533	52718	55255
33597	47071	55618	51796	71027	46690	08002	45066	02870	60012
22828	96380	35883	15910	17211	42358	14056	55438	98148	35384

00631	95925	19324	31497	88118	06283	84596	72091	53987	01477
75722	36478	07634	63114	27164	15467	03983	09141	60562	65725
80577	01771	61510	17099	28731	41426	18853	41523	14914	76661
10524	20900	65463	83680	05005	11611	64426	59065	06758	02892
93815	69446	75253	51915	97839	75427	90685	60352	96288	34248
81867	97119	93446	20862	46591	97677	42704	13718	44975	67145
64649	07689	61711	12169	15238	74106	60655	56289	74166	78561
55768	09210	52439	33355	57884	36791	00853	49969	74814	09270
38080	49460	48137	61589	42724	92035	21766	19435	92579	27683
22360	16332	05343	34613	24013	98831	17157	44089	07366	66196
40521	09057	00239	51284	71556	22605	41293	54854	39736	05113
19292	69862	59951	49644	53486	28244	20714	56030	39292	45166
79504	40078	06838	05509	68581	39400	85615	52314	83202	40313
64138	27983	48048	42631	58658	62243	82572	45211	37060	15017

المصدر: مفاهيم أساسية في الإحصاء. د. محمد المنصوب. منشورات دار الخبرة،



صنعاء، 1998.



10- قائمة المراجع

1.10 المراجع العربية

1- أبو صالح، محمد صبحي.(2001): الطرق الإحصائية . الطبعة الثانية دار البازوري العلمية، عمان: الأردن.

2- البلداوي، عبد الحميد .(1997) : الإحصاء للعلوم الإدارية والتطبيقية . الطبعة العربية الأولى، دار الشروق، عمان : الأردن.

3- توفيق، عبد الجبار. (1983): التحليل الإحصائي في البحوث التربوية والنفسية والاجتماعية الطرق اللامعلمية . الطبعة الأولى، مؤسسة الكويت: الكويت.

4- رمضان، زياد .(1997): مبادئ الإحصاء الوصفي والتطبيقي والحيوي.الطبعة الرابعة، دار وائل للنشر،عمان : الأردن.

5- العواد، منذر حسين .(2003): مبادئ الإحصاء.الطبعة الأولى، مركز الأمين للنشر، صنعاء : اليمن.

6- القاضي، دلال و سهيلةو البياتي، محمود. (2004): الإحصاء للإداريين والاقتصاديين. دار الحامد للنشر والتوزيع، عمان: الإردن.

2- المنصوب، محمد عبد الكريم .(1998): مفاهيم أساسية في الإحصاء . الطبعة الأولى، منشورات دار الخبرة، صنعاء: اليمن.

9- الهيتي، صلاح الدين حسين.(2004): الأساليب الإحصائية في العلوم الإدارية.الطبعة الأولى، دار وائل للطباعة والنشر، عمان: الأردن.

10 . 2 المراجع الأجنبية

1- Dixon & Massey , Introduction to Statistical Analysis.

2 -Hayslett , H . T. (J r) , Statistics Made Simple , Made simple Books , Doubleday and Co ,. Inc . Garden City , New York , 1968.

3- Mood, A.M. (1954): Introduction to the theory of Statistic. New York: McGraw Hill.

11. التعيينات

السؤال الأول :

أختر الإجابة الصحيحة من بين الإجابات الواردة بوضع دائرة حول رقم الإجابة الصحيحة:

1- الإحصاء الوصفي هو الذي يختص بـ:

1- طرق جمع البيانات وتحليلها ووصفها

2- التعامل مع البيانات الإحصائية دون التعميم

3- كل ما ذكر صحيح

2- تنقسم مقاييس المتغيرات النوعية إلى:

1- مقياس ترتيبي ومقياس أسمي 2- مقياس فئوي ومقياس أسمي

3- مقياس نسبي ومقياس فئوي

3- أسلوب الحصر الشامل يقصد به

1- جمع المعلومات عن الظاهرة موضوع الدراسة باستخدام الوحدات المكونة للظاهرة

2- جزء من المجتمع يتم اختياره بطريقة عشوائية لتمثيل ذلك المجتمع

3- ليس أي مما ذكر

4- إن تحديد عدد الفئات المرغوب فيها يتوقف على :

1- حجم البيانات والفروق بين أكبر وأصغر قيمة فيها

2- مستوى الدقة المطلوبة 3- مدى تجانس هذه البيانات

4- كل ما ذكر

5- يهدف استخدام الأعمدة البيانية إلى تحقيق أي من الأهداف التالية:

1- إظهار التطور التاريخي للظاهرة 2- مقارنة ظاهرتين أو أكثر

3- جذب الانتباه إلى اتجاه الأرقام وليس إلى الأرقام ذاتها

4- كل ما ذكر

جامعة العلوم والتكنولوجيا

6- الإحصاء الاستدلالي هو ما يتعلق بـ

1- طرق تحليل وتفسير وتقدير واستخلاص الاستنتاجات

2- هو الذي يتعامل مع التصميم والتقدير والتنبؤ.

3- هو علم اتخاذ القرارات في ظل عدم التأكد

4- كل ما ذكر سابقاً

7- مستوى دخل الفرد هو متغير:

1- نوعي　　　2- كمي　　　3- ليس أي مما ذكر

8- من عيوب أسلوب العينة

1- تحتاج لوقت طويل　　　2- تحتاج لتكاليف عالية

3- تستغرق جهداً أكبر　　　4- ليس أي مما ذكر

9- من الاعتبارات الواجب مراعاتها عند تحديد عدد الفئات

1- أن يكون عدد الفئات بين 5، 15 فئة

2- أن يكون عدد الفئات بين 4، 9 فئات

3- أن يكون عدد الفئات أقل من 15 فئة

10- مركز الفئة (140-144) هو

1- 141　　　2- 142　　　3- 141.5

11- تصنف المتغيرات الإحصائية إلى:

1- متغيرات كمية ومتغيرات نوعية　　　2- متغيرات نوعية ومتغيرات نسبية

3- متغيرات كمية ومتغيرات نسبية

UNIVERSITY OF SCIENCE & TECHNOLOGY
جامعة العلوم والتكنولوجيا

12- يقصد بالمصادر الثانوية للبيانات الإحصائية:

1- الجهات التي تقوم بنفسها بجمع البيانات وتهيئتها ونشرها.

2- الجهات التي تعتمد في جمع البيانات على جهات أخرى وتقوم هي بطبع ونشر البيانات.

3 - الجهات الحكومية التي تقوم بجمع البيانات الإحصائية

4- كل ما ذكر سابقاً

13- تمر عملية تبويب البيانات بثلاث مراحل إحداها:

1- العرض الجدولي

2- تقسيم البيانات إلى مجموعات حسب صفاتها المشتركة

3- توضيح المصدر الذي أخذت منه البيانات الإحصائية

السؤال الثاني:

ضع علامة (√) أمام الجمل الصحيحة وعلامة (X) أمام الجمل الخاطئة فيما يأتي:

1. منظمة الصحة العالمية تعتبر من المصادر الأصلية لجمع البيانات(   ).
2. يتم تحديد عدد الفئات المرغوب فيها بقسمة مدى البيانات على طول الفئة(   ).
3. يمكن وجود فئات مفتوحة في وسط الجدول التكراري. (   )
4. المساحة المحصورة بين المنحنى التكراري والمحور الأفقي لا تساوي المساحة المحصورة بين المضلع التكراري والمحور الأفقي لنفس البيانات(   ).
5. يعتبر الإحصاء الاستدلالي خطوة لاحقه للإحصاء الوصفي. (   )
6. مستوى الدخل متغير نوعي ويقاس بمقياس أسمي(   ) .
7. العامل الأساسي في تحديد حدود الفئة هو نوعية البيانات. (   )
8. يمكن تمثل البيانات الكمية باستخدام الرسوم الدائرية (   ).

2 الوحدة الثانية

المقاييس الإحصائية


UNIVERSITY OF SCIENCE & TECHNOLOGY
١٩٩٤م
١٤١٤هـ
وقل رب
زدني علماً

محتويات الوحدة

1- المقدمــة:

1.1 تمهيـد:

عزيزي الدارس، أرحب بك في مطلع هذه الوحدة (المقاييس الإحصائية) وأذكرك بأننا قد أوضحنا في الوحدة الأولى الأساليب المختلفة لجمع وترتيب وتلخيص وعرض البيانات الإحصائية، وعرفنا أن عرض البيانات باستخدام الرسوم البيانية مفيد للغاية، وأنها تعكس وصفاً عاماً وسريعاً لتلك البيانات مما يتفق مع القول الشائع " بأن صورة واحدة تساوي ألف كلمة " . إلا أن هناك حدود لاستخدام الرسوم البيانية لوصف وتحليل البيانات رغم أهميتها حيث أنها لا تستجيب للأساليب الرياضية اللازمة للتحليل الإحصائي، بالإضافة إلى وجود بعض البيانات التي يصعب عرضها بيانياً، وكذلك صعوبة الاستفادة منها في مجال الاستقراء الإحصائي مما يضطرنا لاستخدام مقاييس وصفية أخرى لإبراز الخصائص الأساسية للبيانات.

تتألف هذه الوحدة من سبعة أقسام رئيسة، تناول القسم الأول منها مقدمة الوحدة. في حين نتناول في القسم الرئيس الثاني منها(مقاييس النزعة المركزية) والتي تستخدم لتحديد مركز البيانات المتاحة وسوف يتبين لك أنواع هذه المقاييس وهي الوسط الحسابي والوسيط والمنوال، وبذلك تنتقل إلى وصف البيانات الإحصائية بدقة أكثر، ويرجع السبب في ذلك لكون هذه المقاييس تعطي قيماً عددية لتحديد المركز الذي تتمحور حوله المفردات والتي من خلالها توصف باقي المفردات من خلال قياس المتوسطات إلا أن هذه المقاييس غير كافية لتحديد خصائص أو صفات التوزيعات التكرارية لأن الوصف يجب أن يكون من خلال أكثر من صفة.

كما تناولت الوحدة في قسمها الثالث مقاييس الموضع (الربيعات والعشيرات والمئينات) وطرق حسابها ليكتمل وصف القيم الموضعية في البيانات.

إن وصف البيانات بشكل كامل يحتم علينا دراسة أكثر من جانب عنها، فما زالت الحاجة ماسة لقياس مدى تقارب أو تباعد مفردات الظاهرة عن بعضها البعض، مما دفعنا إلى استعراض ومناقشة (مقاييس التشتت) كالمدى والانحراف الربيعي والتباين والانحراف المعياري في القسم الرابع من هذه الوحدة بحيث نستطيع وصف البيانات الإحصائية وصفاً كاملاً.

وفي القسم الخامس من هذه الوحدة ناقشنا(مقاييس التشتت النسبية) والتي تستخدم للمقارنة بين مجموعتين أو أكثر مختلفة في وحدات القياس (القيم المعيارية، و معامل الاختلاف).

جامعة العلوم والتكنولوجيا

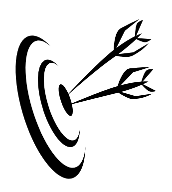
كما تضمن القسم السادس (مقاييس التماثل والالتواء) كيفية حساب معامل الالتواء لتحديد طبيعة التوزيع التكراري بطرق مختلفة وتفسيره دون رسم منحنى التوزيعات التكرارية التي سبق الحديث عنها. ولغاية استكمال وصف التوزيع التكراري للبيانات فقد تم استعراض كيفية حساب معامل التفرطح بطرق مختلفة في القسم السادس.

إضافة إلى ذلك ستجد في ثنايا الوحدة تدريبات متعددة لها حلول نموذجية، تمنحك الفرصة لمراجعة أجزاء الوحدة، كما توجد أسئلة التقويم الذاتي التي تعزز فهمك وتنمي استيعابك.

1 .2 أهداف الوحدة:

بعد فراغك من دراسة هذه الوحدة يجب أن تكون قادراً على أن:

1- توضح مفهوم النزعة المركزية .
2- تذكر مقاييس النزعة المركزية.
3- تحسب مقاييس النزعة المركزية بطريقة صحيحة.
4- تفسر دلالات كل من مقاييس النزعة المركزية بصفتها مقاييس موقع، ومقاييس التشتت بصفتها مقاييس تقارب وتباعد البيانات.
5- تقارن بين المتوسطات الثلاثة (الوسط ، الوسيط ، المنوال)
6- تحسب مقاييس الموضع (الربيعات ، العشيريات ، المئينات).
7- تشرح مفهوم التشتت بنوعيها النسبي والمطلق.
8- تحسب مقاييس التشتت بطريقة صحيحة
9- تحسب وتفسر معامل الاختلاف والقيمة المعيارية.
10- تحدد طبيعة التوزيع التكراري بطرق مختلفة
11- تصف التماثل أو الالتواء أو التفرطح للتوزيعات التكرارية بعد تمثيلها بيانياً.
12- تفرق بين مقاييس النزعة المركزية ومقاييس التشتت.



3.1 أقسام الوحدة

يـرتبط القسـم الأول مـن هـذه الوحـدة ارتباطـاً وثيقـاً بالأهـداف (مـن الأول إلى الخـامس) حيـث أعطينـا في هـذا القسـم نبـذة عـن مفهـوم النزعـة المركـزيـة ودلالات مقـاييس النزعـة المركـزيـة وطـرق حسـابها، والمقارنـة بـين خواصـها. كمـا تناولنـا العلاقة بين المتوسطات الثلاثة (الوسط، و الوسيط، والمنوال).

أما في القسم الثاني فقد تناولنا فيه كيفية حساب مقاييس الموضع (الربيعات، العشيريات، المئينات) وهذا يحقق الهدف السادس.

وتم في القسم الثالث توضيح مفهوم التشتت المطلق والحاجة إلى دراستها كما تم عـرض بعـض مقـاييس التشـتت مثـل المـدى والمـدى الربيعـي والتبـاين والانحـراف المعياري وطرق حسابها وبهذا تتحقق الأهداف الرابع والسابع والثامن.

أما الهدفين الرابع والتاسع فيتحققان بدراستك للقسم الرابع والخاص بمقـاييس التشتت النسبي (معامل الاختلاف والقيمة المعيارية).

بينمـا يـرتبط الهـدفين العاشـر والحـادي عشـر بحسـاب معامـل الالتـواء للتوزيـع التكـراري وتفسـيره وتحديـد طبيعـة التوزيـع، وكـذلك معامـل التفـرطح. ليكتمـل لديك وصف التوزيعات التكرارية، وبهذا يتحقق الهدف الثاني عشر.

1 .4 القراءات المساعدة:



إن مادة هذه الوحدة التي بين يديك تعد كافية لاستيعاب مـا تتضـمنه، ولإثـراء معلوماتـك وترسـيخها حـول موضـوع الوحـدة ننصـحك بـالرجوع إلى القراءات المساعدة الآتية:

1- المنصوب، محمد عبد الكـريم.(1998): مفـاهيم أساسـية في الإحصـاء. الطبعة الأولى، منشورات دار الخبرة ، صنعاء: اليمن.

2- العواد، منذر حسين .(2003): مبـادئ الإحصـاء. الطبعـة الأولى، مركـز الأمين للنشر، صنعاء: اليمن.

3- الهيتي، صـلاح الـدين حسـين.(2004): الأسـاليب الإحصـائية في العلـوم الإدارية.الطبعة الأولى، دار وائل للطباعة والنشر، عمان: الأردن.

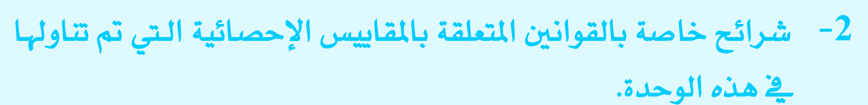
4- أبو صالح، محمد صبحي ومروة أحمد.(2005) : مبادئ الإحصاء. الطبعة الثانية،منشورات جامعة القدس المفتوحة، عمان: الأردن.

1 .5 الوسائط التعليمية المساندة:



قبل أن تبدأ بدراسة هذه الوحدة تأكد من أن حقيبتك التعليمية تحتوي على الوسائط التالية:





1- الآلة الحاسبة

2- شرائح خاصة بالقوانين المتعلقة بالمقاييس الإحصائية التي تم تناولها في هذه الوحدة.

1 .6 ما تحتاج إليه لدراسة الوحدة:

بعد أن تكون قد هيئت المكان المناسب للدراسة، تحتاج إلى توفير المستلزمات الآتية: قلم ، آلة حاسبة وبعض الأدوات للكتابة، وأثناء دراستك المتأنية لمادة الوحدة ستجد في ثنايا هذه الوحدة أسئلة التقويم الذاتي التي تساعدك في مراجعة وتلخيص الوحدة، كما ستجد تدريبات متنوعة عليك القيام بها ، فهي سترسخ الأفكار المطروحة في الوحدة وتمنحك فرصة لاختبار تعلمك ومدى استيعابك لمحتوى هذه الوحدة وفهمك لها.

عزيزي الدارس، وفي حال واجهتك أي صعوبات عليك الاتصال بالمرشد الأكاديمي الذي سيرحب بك ويساعدك على إيجاد الحلول المناسبة لأسئلتك.

2. مقاييس النزعة المركزية Measures of Central Tendency

ذكرنا سابقاً بأن العرض البياني لبيانات التوزيع التكراري يمدنا بصورة مرئية تساعد على فهمها ، إلا أنه على الرغم من أهمية العرض البياني فإنه لا يستجيب للأساليب الرياضية اللازمة لتحليل البيانات إحصائياً ، إلى جانب وجود بعض البيانات التي يصعب عرضها بيانياً ، لذا فإن الإحصائيين لا يكتفون بالعرض البياني فقط بل يفضلون استخدام مقاييس النزعة المركزية بالإضافة إلى بعض المقاييس الإحصائية الأخرى لوصف وتحليل البيانات. فبعد أن ينتهي الباحث من جمع البيانات وتصنيفها وتبويبها وتمثيلها بيانياً ينتقل إلى وصفها عن طريق إبراز الخصائص الأساسية لها والتي يمكن التعبير عنها بمقاييس محددة. والخصائص الأساسية لأي مجموعة من البيانات تقاس بمقاييس معينة ونستعرض معك في هذا القسم مقاييس النزعة المركزية والتي تشتمل على(الوسط الحسابي، الوسيط، المنوال، الوسط التوافقي، الوسط الهندسي) ولا يمكن تفضيل أحد هذه المقاييس على الأخرى لأن لكل منها مزاياه وعيوبه، إلا أن هناك بعض التوزيعات يصلح فيها استخدام أحد هذه المقاييس أكثر من الآخر.

1.2 مفهوم النزعة المركزية (The Central Tendency Concept):

هناك ميل لأن تتجمع مفردات الكثير من التوزيعات حول قيمة معينة في التوزيع وهذا الميل يسمى بالنزعة المركزية أي نزعة المفردات المختلفة للتجمع حول مركز معين. وتستخدم مقاييس النزعة المركزية لتحديد مركز البيانات المتاحة. ويجب ملاحظة أن وصف كيفية انتشار البيانات على جانبي القيمة المتوسطة لا يقل أهمية عن تحديد مركز البيانات.

وتُعرّف النزعة المركزية بأنها: ميل معظم المفردات المختلفة للتجمع حول نقطة أو قيمة واحدة تسمى القيمة المتوسطة. فالقيمة المتوسطة لمجموعة ما من القيم هي قيمة نموذجية يتم اختيارها لتكون دليلاً مميزاً وممثلاً لقيم المجموعة.

UNIVERSITY OF SCIENCE & TECHNOLOGY ١٩٩٤م وقل رب زدني علما جامعة العلوم والتكنولوجيا

2.2 الوسط الحسابي (The Arithmetic Mean):

تستخدم كلمة متوسط حسابي في الحياة اليومية كثيراً .فنقول مثلاً أن درجة الطالب خالد أعلى من المتوسط عندما نرد على سؤال بشأن تحصيله الدراسي أو نقول متوسط الدخل للفرد أو متوسط الثقافة أو متوسط الاستهلاك اليومي للقهوة...الخ هذه الاستخدامات الشائعة لكلمة متوسط تعبر عن شعور داخلي معين يحسه ويفهمه كل منا ولا يستطيع ترجمته بدقة ، ومقاييس النزعة المركزية ومنها المتوسط الحسابي هي محاولة لترجمة هذا الشعور بطريقة محددة ودقيقة تماماً.

ويعرف الوسط الحسابي بأنه: مجموع مفردات المجموعة مقسوماً على عددها.

و يعرف بأنه: ذلك المقياس الإحصائي الوصفي الذي إذا حسبنا انحرافات مفردات المجموعة عنه لكان مجموع هذه الانحرافات يساوي الصفر.

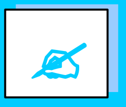
1.2.2 طرق حساب الوسط الحسابي :

1. في حالة البيانات غير المبوبة (Ungrouped Data):

إذا كانت لدينا مجموعة قيم $x_1, x_2, ..., x_n$ فإن وسطها الحسابي والذي يرمز له بالرمز $\bar{X}$ (تقرأ اكس بار) هو خارج قسمة مجموعها على عددها

$$\bar{X} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{x_1 + x_2 + ... + x_n}{n}$$

والرمز $\sum$ (سيجما) يعني جمع جميع الحدود في دالة



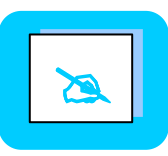
مثال(1.2) :

أوجد الوسط الحسابي للمفردات التالية 8،13،16،7،6،25،9

الحل

$$\because \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$\therefore \bar{x} = \frac{8+13+16+7+6+25+9}{7} = 12$$



مثال (2.2):

إذا كان الأجر الشهري لمجموعة من عمال أحد المصانع كالآتي:

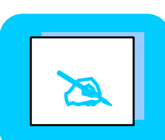
3000، 4000، 4500، 4000، 3500، 3000، 3000، 4000

المطلوب: أوجد الوسط الحسابي لهذه الأجور.

الحل

$$\bar{x} = \frac{3000+4000+4500+4000+3500+3000+3000+4000}{8}$$

$$\bar{x} = \frac{29000}{8} = 3625$$



مثال(3.2):

احسب الوسط الحسابي للبيانات التالية:

25، 29، 13، 17، 32، 27، 19

الحل:

$$\bar{x} = \frac{25+29+13+17+32+27+19}{7} = 23.1429$$

2. في حالة البيانات المبوبة (Grouped Data):

هناك عدة طرق لحساب الوسط الحسابي لجدول تكراري وهي:

1- الطريقة المباشرة.

2- طريقة الوسط الفرضي.

3- طريقة الانحرافات المختصرة.

وهذه الطرق تنطلق من افتراض أن مركز الفئة في الجدول التكراري هو مقياس وصفي لكل تكرار في هذه الفئة أي أن مركز الفئة يمثل جميع مفردات تلك الفئة.

أ- الطريقة المباشرة

يتم حساب الوسط الحسابي للجداول التكرارية بالطريقة المباشرة حسب الخطوات الآتية:

1- نكتب البيانات في شكل فئات متساوية أو غير متساوية.

2- نوجد مراكز الفئات (x_i) .

3- نوجد حاصل ضرب مراكز الفئات في تكراراتها أي: $(f_i x_i)$

4- نحسب الوسط الحسابي باستخدام العلاقة .

$$\overline{X} = \frac{\sum_{i=1}^{n} f_i X_i}{\sum_{i=1}^{n} f_i}$$

حيث : $\overline{X}$: هي الوسط الحسابي

X_i : هي مركز الفئة

f_i : هي التكرارات

مثال (4.2):



احسب الوسط الحسابي للبيانات الموضحة بالتوزيع التكراري الآتي:

الفئات	10 – 14	15 – 19	20 – 24	25 – 29	30– 34	35 – 39
التكرارات f_i	2	6	8	5	5	4

الحل:

جدول(1.2)

الوسط الحسابي بالطريقة المباشرة

فئات	f_i	X	$f_i x_i$
10 – 14	2	12	24
15 – 19	6	17	102
20 – 24	8	22	176
25 – 29	5	27	135
30– 34	5	32	160
35 – 39	4	37	148
$\sum$	30	---	745

$$\because \ \overline{x} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i} \quad \Rightarrow \quad \therefore \ \overline{x} = \frac{745}{30} = 24.83$$

ومن الملاحظ أن هذه الطريقة طويلة وبصفة خاصة إذا كان الجدول كبير والتكرارات كبيرة ويفضل استخدام طرق أخرى أسهل من حيث إجراء العمليات الحسابية.

ب- طريقة الوسط الفرضي:

تستخدم هذه الطريقة إذا كانت الأرقام التي سيتم معالجتها كبيرة بهدف تبسيط إجراء العمليات الحسابية حسب الخطوات الآتية:

1- تحديد مراكز الفئات وتكراراتها

2- تحديد وسط فرضي من بين مراكز الفئات ويفضل أن يكون مركز الفئة ذات التكرار الأكبر.

3- نحسب انحراف مراكز الفئات عن الوسط الفرضي وذلك بطرح الوسط الفرضي من مركز كل فئة ووضع النتائج في عمود جديد.

4- نوجد حاصل ضرب كل انحراف بتكرار الفئة التي يعود إليها ونضع النتائج في عمود أخر.

5- يحسب الوسط الحسابي من العلاقة التالية:

$$\overline{X} = L + \frac{\sum f_i \ d_i}{\sum f_i}$$

حيث L : الوسط الفرضي

d : انحراف مركز الفئة عن الوسط الفرضي

<u>**مثال(5.2):**</u>



من بيانات المثال(4.2) احسب الوسط الحسابي باستخدام طريقة الوسط الفرضي

الحل:

نلاحظ أن أكبر تكرار هو 8 وهو تكرار الفئة 24 – 20 وبذلك يمكن تحديد مركز هذه الفئة (22) ليكون وسطاً فرضياً ونتابع الخطوات كما يلي:

جامعة العلوم والتكنولوجيا
UNIVERSITY SCIENCE & TECHNOLOGY

جدول (2.2)

الوسط الحسابي بطريقة الوسط الفرضي

الفئات	التكرار f_i	مراكز الفئات x_i	الانحراف d_i	الانحراف × التكرار $f_i d_i$
10 – 14	2	12	- 10	-20
15 – 19	6	17	-5	-30
20 – 24	8	22	0	0
25 – 29	5	27	5	25
30– 34	5	32	10	50
35 – 39	4	37	15	60
المجموع	30	--	--	85

الوسط الفرضي

$$\bar{X} = L + \frac{\sum f_i d_i}{\sum f_i}$$

$$\bar{X} = 22 + \frac{85}{30} = 24.83$$

ج- طريقة الانحرافات المختصرة:

إن الهدف من استخدام هذه الطريقة هو تبسيط العمليات الحسابية ويفضل استخدام هذه الطريقة في حالة الجداول المنتظمة وتتم العمليات الحسابية حسب الخطوات الآتية:

1- تحديد مراكز الفئات وتكراراتها

2- تحديد وسط فرضي من بين مراكز الفئات

3- نحسب انحراف مراكز الفئات عن الوسط الفرضي وذلك بطرح الوسط الفرضي من مركز كل فئة وذلك بوضع النتائج في عمود جديد.

4- نقسم كل قيمة في عمود الانحرافات على طول الفئة (c) [1] فنحصل على الانحرافات المختصرة (d_m).

---

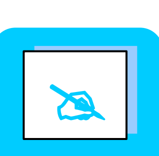
[1] - إذا كانت أطوال الفئات غير متساوية تتم القسمة على أصغر طول فئة

5- نوجد حاصل ضرب كل انحراف في عمود الانحرافات المختصرة بتكرار الفئات التي يعود إليها ونضع النتائج في عمود أخر. ويحسب الوسط الحسابي من العلاقة الآتية:

$$\bar{X} = L + \frac{\sum f_i d_m}{\sum f_i} . c$$

حيث L : الوسط الفرضي و d_m : الانحرافات المختصرة و c : طول الفئة

مثال(6.2):



من بيانات المثال (4.2) احسب الوسط الحسابي بطريقة الانحرافات المختصرة.

الحل:

امتداداً لمعطيات الجدول (1.2) فإن $c = 5$ طول الفئة ، ثم نستكمل بقية الحل كما في الجدول (3.2).

جدول(3.2)

الوسط الحسابي بطريقة الانحرافات المختصرة

الفئات	f_i	X_i	d	$d_m = \frac{d}{c}$	$(d_m)(f_i)$
10 – 14	2	12	-10	-2	-4
15 – 19	6	17	-5	-1	-6
20 – 24	8	22	0	0	0
25 – 29	5	27	5	1	5
30– 34	5	32	10	2	10
35 – 39	4	37	15	3	12
المجموع	30				17

ومن العلاقة $\bar{X} = L + \frac{\sum f_i d_m}{\sum f_i} . c$ نجد أن

$$\bar{X} = 22 + \frac{17}{30} .5 = 24.83$$

لا حظ **عزيزي الدارس** أن قيمة الوسط الحسابي بالطرق الثلاث لم يختلف، حيث كان الهدف هو تبسيط إجراء العمليات الحسابية.

2.2.2 بعض خواص الوسط الحسابي

1- مجموع انحرافات القيم عن وسطها الحسابي يساوي صفر أي أن

$$\sum_{i=1}^{n}(x_i - \bar{x}) = 0$$

2- تتأثر قيمة الوسط الحسابي كثيراً بالقيم المتطرفة (القيم الصغيرة جداً أو الكبيرة جداً)

3 - يأخذ جميع القيم بعين الاعتبار.

4- الوسط الحسابي محدد جبرياً بدقة ويتمتع بخواص جبرية لا يتمتع بها غيره من المقاييس.

5- لا يمكن حساب الوسط الحسابي للبيانات النوعية.

6- لا يمكن حساب الوسط الحسابي من الجداول التكرارية المفتوحة.

7- لا يمكن إيجاد الوسط الحسابي بالطريقة البيانية إلا في حالة التوزيع التكراري المتماثل.

8- لتكن $x_1 , x_2 , x_3 ...., x_n$ مجموعة من القيم، وسطها الحسابي $\bar{X}$ فإذا أضفنا لكل قيمة من القيم الأصلية عدداً ثابتاً C فإن الوسط الحسابي الجديد يساوي الوسط الحسابي للقيم الأصلية مضافاً إليه نفس المقدار الثابت أي $\bar{Y}$ يصبح $\bar{X} + C$. وبالمثل في حالة طرح مقدار ثابت من جميع القيم فإن الوسط الحسابي الجديد $\bar{Y}$ يكون على الصورة : $\bar{Y} = \bar{x} - c$

9- لتكن $x_1 , x_2 , x_3 ...., x_n$ مجموعة من القيم، وسطها الحسابي $\bar{X}$ فإذا ضربنا كل قيمة من القيم الأصلية بعدد ثابت C فإن الوسط الحسابي الجديد $\bar{Y}$ يساوي الوسط الحسابي للقيم الأصلية مضروبا في المقدار الثابت نفسه $\bar{Y} = c\,\bar{x}$ وبالمثل نجد $\bar{Y} = \frac{\bar{x}}{c}$ في حالة قسمة كل مفردة على عدد ثابت.

لا يستخدم الوسط الحسابي في الحالات الآتية:
- عندما يكون التوزيع شديد الالتواء.
- عندما يكون هناك معدل نمو أو تغير حول فترة زمنية.
- عند حساب معدلات مثل السرعة والأسعار..إلخ.
- عندما تأخذ البيانات شكل التتابع الهندسي مثل: 1، 2، 3، 4، ....



مثال (7.2):

إذا كانت البيانات التالية تمثل أعمار خمسة أشخاص بالسنوات 22، 32، 18، 20، 24 فأوجد الوسط الحسابي لهذه الأعمار.

الحل:

$$\because \bar{X} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{22+31+18+20+24}{5}$$

$$\therefore \bar{X} = \frac{115}{5} = 23$$

مثال(8.2):

لو أضفنا العدد 5 إلى كل من أعمار الأشخاص الواردة في مثال(7.2) فإن متوسط الأعمار يصبح



$$\because \bar{X} = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$\therefore \bar{x} = \frac{(22+5)+(31+5)+(18+5)+(20+5)+(24+5)}{5}$$

$$= \frac{140}{5} = 28$$

أي أن الوسط الحسابي الجديد يصبح $\bar{x} + c$ حيث $c = 5$ أي

$$23 + 5 = 28$$

مثال (9.2):

لو ضربنا كل من أعمار الأشخاص الواردة في مثال(7.2) بالمقدار الثابت 5 فإن متوسط الأعمار يصبح.



الحل:

$$\because \bar{X} = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$\therefore \bar{X} = \frac{5.(22)+5.(31)+5.(18)+5.(20)+5.(24)}{5}$$

$$= \frac{575}{5} = 115$$

أي أن الوسط الحسابي الجديد يصبح $c.\bar{X}$ حيث $c = 5$ أي

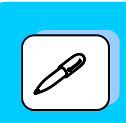
$5.(23) = 115$

تدريب (1)



1. اذكر مقاييس النزعة المركزية؟
2. عرف الوسط الحسابي.

تدريب (2)



أوجد الوسط الحسابي للبيانات الآتية :

20 ، 18 ، 12 ، 10 ، 14 ، 22 ، 23 ، 9 ، 25

تدريب (3)



-أذكر بعض خواص الوسط الحسابي

-وما أهم عيوبه.

أسئلة التقويم الذاتي (1) :

1) أوجد الوسط الحسابي للتوزيع التكراري الآتي:

مركز الفئة	التكرار
45	10
55	15
65	22
75	13
85	7

2) متوسط 23 فقرة هو 17.3 فما هو مجموع هذه القيم.

3) أوجد الوسط الحسابي للبيانات الآتية:

0.001 ، 0.003 ، 0.002 ، 0.004

نلاحظ من خلال دراسة الوسط الحسابي أنه يتأثر بالقيم المتطرفة مما يفقد المتوسط الموقع المركزي الذي من المفترض أن يشغله بالإضافة إلى أن بعض البيانات تكون وصفية أو تراجع ولا يوجد معنى لكلمة متوسط في مثل هذه البيانات. وسوف نتطرق هنا إلى مقياس آخر من مقاييس النزعة المركزية لا يتأثر بالقيم المتطرفة وهذا المقياس هو الوسيط.

3.2 الوسيط (The Median):

تعريف: الوسيط هو القيمة التي تتوسط مجموعة القيم بحيث يسبقها نصف عدد القيم ويتلوها النصف الآخر. ويمكن الحصول على الوسيط بأن ترتب درجات المجموعة ترتيباً تنازلياً أو تصاعدياً ثم نأخذ القيمة التي تقع في المنتصف تماماً إذا كان عدد الدرجات فردياً أما إذا كان عدد الدرجات زوجياً فإن قيمة الوسيط تساوي الوسط الحسابي للقيمتين الواقعتين في الوسط..

يعتبر الوسيط أقل وثوقاً من الوسط الحسابي في مسائل الاستقراء (التقدير، التنبؤ، ....إلخ). وهو أهم ما يميز الوسط الحسابي عن الوسيط.

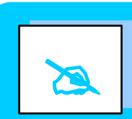
1.3.2 طرق حساب الوسيط:

1- في حالة البيانات غير المبوبة (Ungrouped Data)

أ- إذا كان عدد مفردات المجموعة فردياً فإن الوسيط هو القيمة التي يدل عليها ترتيب القيمة الوسيطية في تلك المجموعة ويحسب ترتيب الوسيط من العلاقة:

$$\text{ترتيب الوسيط} = \frac{\text{عدد الدرجات} + 1}{2}$$

<u>مثال(10.2):</u>



أوجد الوسيط للأعداد التالية: 5، 4، 3، 8، 7

الحل :

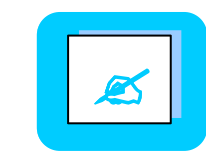
نرتب الأعداد ترتيباً تصاعدياً 3، 4، 5، 7، 8

بما أن عدد القيم فردياً نوجد ترتيب القيمة الوسيطية كما يلي:

$$\frac{n+1}{2} \quad \text{أي أن} \quad 3 = \frac{5+1}{2} = \text{رتبة الوسيط}$$

أي أن القيمة الوسيطية هي القيمة ذات الرتبة الثالثة وبالتالي فإن الوسيط هو القيمة 5.

مثال(11.2):



في فصل دراسي يتضمن 9 طلاب كانت التقديرات في أحد الاختبارات كما يلي: جيد، ضعيف، مقبول، جيد، جيد جداً، ممتاز، مقبول، جيد، جيد جداً. احسب الوسيط .

الحل :

نرتب التقديرات تصاعدياً كما يلي:

ضعيف، مقبول، مقبول، جيد، جيد، جيد، جيد جداً، جيد جداً، ممتاز.

نلاحظ أن الوسيط هو القيمة ذات الترتيب الخامس لأن عدد المفردات فردياً أي :

$$\text{ترتيب الوسيط} = \frac{9+1}{2} = 5$$

وعليه يكون الوسيط هو التقدير " جيد"

ب- إذا كان عدد مفردات المجموعة زوجياً فإن الوسيط هو القيمة الواقعة بين القيمتين اللتين تقعان في وسط المجموعة وبناء على ذلك نقوم بتحديد ترتيب الوسيط الأول والثاني.

$$\text{ترتيب الوسيط الأول} = \frac{\text{عدد الدرجات}}{2}$$

$$\text{ترتيب الوسيط الثاني} = \frac{\text{عدد الدرجات}}{2} + 1$$

$$\text{قيمة الوسيط} = \frac{\text{قمة الوسيط الأول + قيمة الوسيط الثاني}}{2}$$

مثال (12.2):



احسب الوسيط للأعداد الآتية: 13، 12، 5، 8، 6، 9

الحل:

نرتب الأعداد تنازلياً 13، 12، 9، 8، 6، 5 وبما أن عدد المفردات زوجياً فإن الوسيط هو الوسط الحسابي للعددين الثالث والرابع (8 و 9) ترتيب الوسيط الأول = 6 /2 =3 وقيمة الوسيط الأول =9

ترتيب الوسيط الثاني=(6 /2)+1= 4 وقيمة الوسيط الثاني = 8 وبما أن

$$\text{قيمة الوسيط} = \frac{\text{قمة الوسيط الأول + قيمة الوسيط الثاني}}{2}$$

إذا قيمة الوسيط = 2/(8+9) = 8.5.

2- في حالة البيانات المبوبة (التوزيعات التكرارية) (Grouped Data)

يتم حساب الوسيط في حالة البيانات المبوبة بإتباع الخطوات الآتية:

- تصميم جدول توزيع تكراري صاعد
- تحديد موقع الوسيط بقسمة مجموع التكرارات على 2 $(\frac{\sum f_i}{2})$
- تحديد الفئة الوسيطية فإذا كانت قيمة موقع الوسيط تساوي تكرار أي فئة فإنها تكون الفئة الوسيطية . وإذا وقعت بين قيمتين فإن الفئة اللاحقة لقيمة الوسيط تكون هي الفئة الوسيطية.
- نحسب قيمة الوسيط من العلاقة:

$$M_d = L + \frac{\frac{\sum f_i}{2} - f_1}{f_2} * H$$

حيث: L : الحد الأدنى للفئة الوسيطية

f_1 : التكرار المتجمع الصاعد السابق للفئة قبل الوسيطية.

f_2 : تكرار الفئة الوسيطة

H: طول الفئة.

مثال(13.2):



الجدول التالي يوضح بيانات عن الدخل الشهري لـ 100 عامل من عمال المهن الحرة احسب الوسيط لهذا التوزيع التكراري.

فئات الدخل الشهري بالألف	5-	10-	15-	20-	25-	30-	35-	40-	45 – 50
التكرار f_i	10	4	12	20	10	8	20	6	10

الحل:

جدول(4.2)

يوضح بيانات عن الدخل الشهري لـ 100 عامل من عمال المهن الحرة.

فئات الدخل الشهري	f_i	تكرار متجمع صاعد
5 - 10	10	10
10 - 15	4	14
15 - 20	12	26
20 - 25	20	46
25 - 30	10	56
30 - 35	8	64
35 - 40	20	84
40 - 45	6	90
45 - 50	10	100
المجموع	100	--

نحدد موقع الوسيط من خلال العلاقة $\frac{\sum_{i=1}^{k} f_i}{2} = \frac{100}{2} = 50$

وبعد تحديد موقع الوسيط نلاحظ أن القيمة 50 موجودة ضمن الـ 56 في التكرارات المتجمعة الصاعدة، وذلك مقابل الفئة 30 –25 وعليه فإننا على يقين أن قيمة الوسيط لا يمكن بحال من الأحوال أن تقل عن 25 أو تزيد عن 30، وبناء عليه، نستخدم الصيغة التي سقناها آنفاً لحساب قيمة الوسيط كما يأتي:

من الجدول (4.2) نحصل على القيم كما يأتي:

$h = 5$ و $f_2 = 10$ و $f_1 = 46$ و $L = 25$

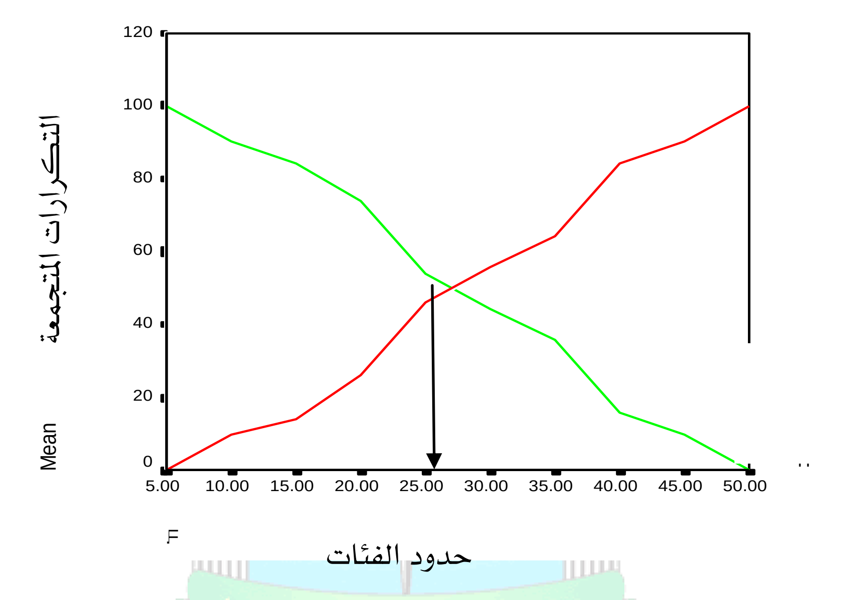
$$M_d = 25 + \left(\frac{50 - 46}{10}\right) \cdot 5 = 27$$

2.3.2 تحديد الوسيط بيانياً

يـتم الحصـول علـى الوسـيط مـن خـلال رسـم المنحـنيين المـتجمعين الصـاعد والهابط، ومن ثم إنزال خط عمودي من نقطة تقاطع المنحنيين على المحور الأفقي، وتلك النقطة تمثل قيمة الوسيط كما هو مبين في الشكل (1.2)

شكل (1.2)

يوضح كيفية تحديد قيمة الوسيط بيانياً من بيانات الجدول(4.2)

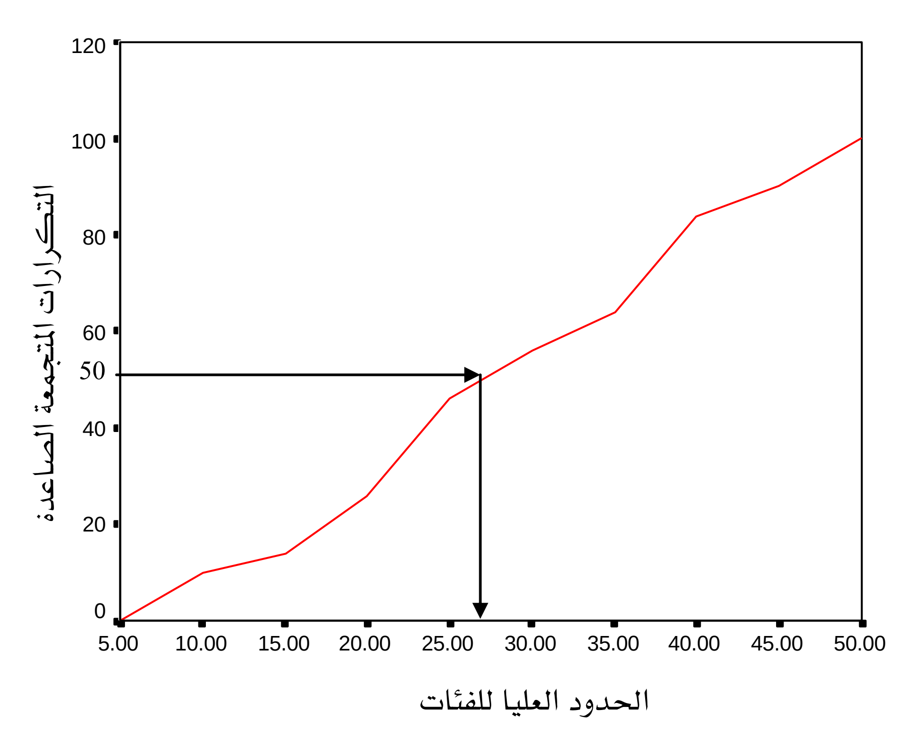


من خلال الشكل لاحظ **عزيزي الدارس** أن قيمة الوسيط $M_d = 27$ أي أنه لا فـرق بـين النتـائج الـتي نحصـل عليهـا بالرسـم البيـاني، وتلـك الـتي نحصـل عليهـا بالحساب، وإذا وجد اختلاف بسيط فذلك يرجع إلى عدم الدقة في الرسم .

كمـا قـد يـتم الاكتفـاء برسـم أحـد المنحيين، المتجمـع الصـاعد أو المتجمـع الهابط ثم نقوم برسم خط مستقيم من موقع الوسيط (والـذي يسـاوي نصـف مجمـوع التكـرارت) يـوازي محـور السـينات ومـن نقطـة تقـاطع هـذا المسـتقيم مـع المنحنـى التكـراري المتجمـع نسـقط عمـوداً علـى محـور السـينات فيتقـاطع معـه عنـد قيمـة الوسيط حسب الشكل(2.2)

شكل(2.2)

يوضح طريقة حساب الوسيط بيانياً



نلاحظ أن قيمة الوسيط = 27

جامعة العلوم والتكنولوجيا

3.3.2 خواص الوسيط

1- يقع الوسيط في أي توزيع تكراري غير متماثل بين الوسط الحسابي والمنوال

2- يتأثر الوسيط بالقيم الوسطى ولا يتأثر بالقيم المتطرفة.

3- سهولة حساب قيمته

4- يمكن حسابه إذا كانت إحدى الفئات مفتوحة من أعلى أو من أسفل الجدول، ويمكن حسابه إذا كانت الفئات غير متساوية الطول

5- مقياس غير محدد بدقة في الجداول التكرارية إذ أن طريقة حسابه تعطي قيمة تقريبية

6- لا يتمتع بأي خواص جبرية بحيث يمكن الاستفادة منه في حسابات لاحقة.

تدريب (4)



احسب الوسيط للأعداد الآتية:

22 , 25 , 14 , 11 , 35 28

تدريب (5)



1- أذكر أهم خصائص الوسيط؟

2- كيف يتم تحديد موقع الوسيط؟

أسئلة التقويم الذاتي (2) :



1) أحسب الوسيط للبيانات الآتية:

2 ، 2 ، 3 ، 7، 10 ، 100

2) جدول التوزيع التكراري الآتي يلخص مجموعة بيانات عن درجة تلوث المياه في ثلاث وثلاثون منطقة سكنية.

الفئة	10-19	20-29	30-39	40-49	50-59	60-69
التكرار	5	7	4	8	4	5

أحسب الوسيط لهذه البيانات.

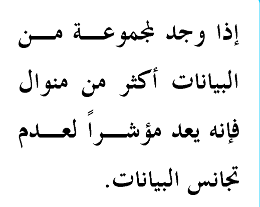
4.2 المنوال (The Mode):

المنوال لمجموعة من القيم هو : القيمة الأكثر تكراراً بين مجموعة القيم، أو هو القيمة التي لها أكثر تكراراً في التوزيعات التكرارية البسيطة. ويمكن

استخدامه للقيم النوعية والكمية، وهو أقل مقاييس النزعة المركزية استعمالا ونرمز له بالرمز M_0.

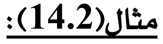
4.2. طرق حساب المنوال :

يمكن حساب المنوال بطرق مباشرة أو غير مباشرة ومنها:



أ- **حساب المنوال من البيانات غير المبوبة**

نستعرض البيانات غير المبوبة ونؤشر على المفردات التي يتكرر حدوثها حيث تكون المفردة التي يتكرر حدوثها أكثر من غيرها هي المنوال وطبقاً لذلك فإن قيمة المنوال قد لا تكون وحيدة فقد تكون البيانات وحيدة المنوال أو ثنائية المنوال أو ثلاثية المنوال إذا تكرر حدوث مفردتين أو أكثر من غيرهما بعدد المرات نفسها، وقد لا يوجد للبيانات منوال وذلك عندما لا تتكرر مفردة أكثر من غيرها.



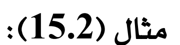


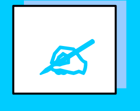
مثال(14.2):

أوجد المنوال للقيم الآتية : 5،4،5،5،3،8، 8،9،6

الحل:

نلاحظ أن القيمة 5 تكررت 3 مرات بينما بقية القيم تكررت عدد قل من المرات فتكون قيمة المنوال هي 5.



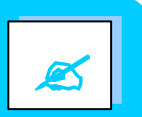


مثال (15.2):

إذا كانت أوزان مجموعة من طلاب المستوى الأول كما يلي:
50، 55، 63، 49، 63، 55، 60 ، 63 فأوجد المنوال لهذه الأوزان.

الحل:

نلاحظ أن الوزن 63 تكرر ثلاث مرات، فيكون منوال هذه الأوزان هو 63



مثال(16.2):

القيم التالية تمثل درجات 6 طلاب في أحد الاختبارات لمادة المحاسبة 85، 69، 88، 92، 91، 75 أوجد منوال هذه الدرجات.

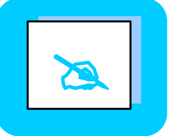
الحل :

من خلال ملاحظة الدرجات الواردة في المثال نجد أنه لا يوجد أي عدد قد تكرر وعليه فإنه لا يوجد منوال لهذه البيانات.

ب- حساب المنوال من البيانات المبوبة ذات التوزيع التكراري البسيط

إذا كان التوزيع التكراري بسيط ليس ذو فئات فإن المنوال هو القيمة التي لها أكبر تكرار.

مثال (17.2) :



احسب المنوال للتوزيع التكراري التالي

x_i	8	7	6	5	4	3	2	5
f_i	2	3	14	6	5	8	7	9

الحل:

تبين من الجدول السابق أن أكثر الأرقام تكراراً هو الرقم 6 لأنه يقابله أكبر تكرار 14 .

إذاً المنوال = 6

جـ- حساب المنوال من التوزيعات التكرارية

عند وضع البيانات في توزيع تكراري للقيم فإننا نلاحظ أن هناك فئة أو أكثر لها تكرار أكبر من بقية الفئات تدعى بالفئة المنوالية وهي الفئة التي يقابلها أكبر تكرار. ويمكن تعيين الفئة المنوالية بمجرد النظر في الجدول التكراري إذا كانت فئات التوزيع متساوية الطول . أما إذا كانت الفئات غير متساوية الطول فإنه يجب تعديل التكرارات قبل تعيين الفئة المنوالية. ولحساب قيمة المنوال يمكن استخدام عدداً من الطرق أهمها طريقة الفروق(طريقة بيرسون) ، ويحسب المنوال بموجب هذه الطريقة من الصيغة:

$$M_0 = L + \frac{d_1}{d_1 + d_2} * H$$

حيث : M_0 : المنوال

L : الحد الأدنى للفئة المنوالية

d_1 : الفرق بين تكرار الفئة المنوالية وتكرار الفئة السابقة للفئة المنوالية.

d_2 :الفرق بين تكرار الفئة المنوالية وتكرار الفئة اللاحقة للفئة المنوالية.

H : طول الفئة المنوالية.

مثال(18.2):

أوجد المنوال من بيانات الجدول التكراري الآتي:

الفئات	20-	25-	30-	35-	40-	45-	50-	55-	60 – 65
التكرار	3	9	13	16	20	15	13	8	3

الحل:

من الجدول نلاحظ أن الفئة المقابلة لأكبر تكرار هي 45 – 40 وأن تكرار الفئة قبل المنوالية هو 16 وتكرار الفئة بعد المنوالية هو 15 وطول الفئة المنوالية = 5

وبما أن:

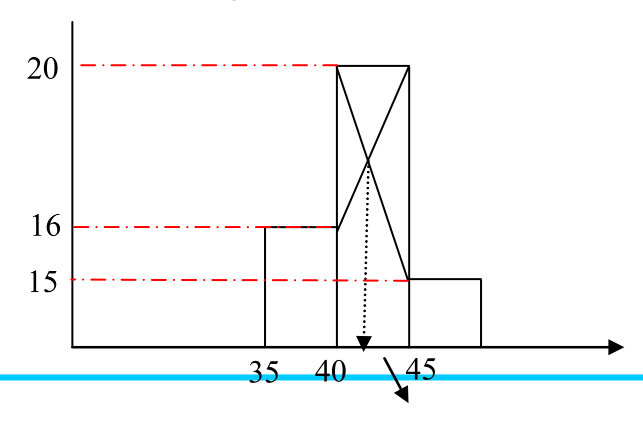
$$M_0 = L + \frac{d_1}{d_1 + d_2} \cdot H$$

$$M_0 = 40 + \frac{4}{4+5} \cdot 5 = 42.22$$

2.4.2 تحديد المنوال بيانياً:

يمكن تحديد قيمة المنوال بيانياً باستخدام المستطيلات التي تمثل تكرار الفئة قبل المنوالية وتكرار الفئة بعد المنوالية حيث نصل الزاوية اليمنى العليا للمستطيل الذي يمثل تكرار الفئة المنوالية بالزاوية التي تماثلها في مستطيل تكرار الفئة قبل المنوالية، ونصل الزاوية اليسرى العليا للمستطيل الذي يمثل تكرار الفئة المنوالية بالزاوية التي تماثلها في مستطيل تكرار الفئة بعد المنوالية فيتقاطع المستقيمان في نقطة داخل المستطيل الذي يمثل تكرار الفئة المنوالية. نسقط عموداً من نقطة التقاطع على محور السينات فيقطعه في نقطة($b \approx 42$)التي تدل على قيمة المنوال كما في الشكل (3.2) والخاص ببيانات المثال (18.2) .

شكل(3.2) يوضح المنوال بيانياً



قيمة المنوال تقريباً = 42

3.4.2 خواص المنوال

1-لا يتأثر المنوال بالقيم المتطرفة والوسطى في التوزيع التكراري بل يتأثر بالتكرارات

2- يتأثر بعدد فئات التوزيع التكراري ومداها .

3- يمكن حسابه من الجداول التكرارية المفتوحة

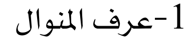
4- ليس مناسباً للعمليات الجبرية

5- يعتمد على جزء من البيانات

6- قد يوجد أكثر من منوال للبيانات وفي هذه الحالة يقال أن البيانات غير متجانسة

7- يستخدم في حالة البيانات النوعية

تدريب (6)



1-عرف المنوال

2-اذكر أهم خواص المنوال

3- حدد منوال البيانات الآتية:

105 , 110 , 105 , 120 , 120 , 105

تدريب (7)

أوجد المنوال للبيانات التالية 17.6 ، 13.5 ، 21.4 ، 11.4، 13.5 ، 19.8، 13.5 ، 17.6



5.2 الوسط التوافقي (Harmonic mean):

يستخدم عند إيجاد المتوسط وفقاً لوحدة قياس معينة ويرمز له بالرمز $\bar{x}_h$ وهو عبارة عن مقلوب الوسط الحسابي لمقلوب القيم.

طرق حسابه :

أ. **من بيانات غير مبوبة :**

$$\bar{X}_h = \frac{n}{\sum_{i=1}^{n} \frac{1}{X_i}}$$

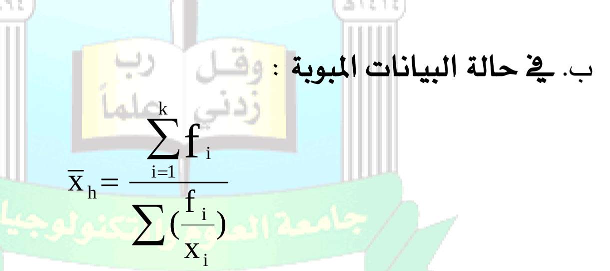


مثال (19.2):

أوجد الوسط التوافقي للقيم : 3،5،6،6،7،10،12

الحل:

$$\bar{x}_h = \frac{7}{\frac{1}{3}+\frac{1}{5}+\frac{1}{6}+\frac{1}{6}+\frac{1}{7}+\frac{1}{10}+\frac{1}{12}} = \frac{7}{\frac{501}{420}} = 5.87$$



ب. **في حالة البيانات المبوبة :**

$$\bar{x}_h = \frac{\sum_{i=1}^{k} f_i}{\sum (\frac{f_i}{x_i})}$$

ومن أبرز مزاياه أنه :

1- يعتمد على جميع القراءات.

2- قابل للمعالجات الرياضية.

6.2 الوسط الهندسي The Geometric Mean

إن تنوع وتعدد مصادر البيانات الإحصائية تجعلنا في حاجة للتعامل مع بيانات ذات طبيعة متغيرة عبر الزمن، كالأسعار ومعدل النمو والأرباح المركبة......الخ وبالتالي نحتاج إلى إيجاد معدل أو متوسط مناسب لتلك المتغيرات. فاستخدام الوسط

الحسابي قد لا يكون له معنى في هذه الحالات لذلك نلجأ إلى مقياس آخر هو ما يعرف بالوسط الهندسي.

تعريف: لتكن $x_1 , x_2 , x_3 ...., x_n$ مجموعة من القيم، فإن وسطها الهندسي هو الجذر النوني لحاصل ضربها. أي يمكن حسابه للبيانات غير المبوبة من العلاقة:

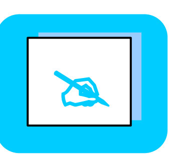
$$G = \sqrt[n]{x_1 . x_2 . ...x_n}$$

وعادة ما نستخدم خصائص اللوغاريتمات لحساب الوسط الهندسي فتأخذ المعادلة السابقة الصورة الآتية :

$$\log G = \frac{\log x_1 + \log x_2 + ...+\log x_n}{n} = \frac{\sum_{i=1}^{n} \log x_i}{n}$$

ثم نأتي بمقلوب اللوغاريتم للحصول على G .

مثال(20.2):



احسب الوسط الهندسي للقيم 100، 5 ، 2

الحل:

$$G = \sqrt[3]{2 \ x \ 5 \ x \ 100} = \sqrt[3]{1000} = 10$$

من مزايا الوسط الهندسي:

1- يعتمد على جميع القيم .

2- محدد جبرياً بدقة

3- قليل التأثر بالقيم المتطرفة نحو اليمين بعكس الوسط الحسابي . لذلك إذا كان التوزيع ملتوي نحو اليمين فإن الوسط الهندسي يكون أفضل المقاييس لتمثيل البيانات.

4- يعتبر ذا أهمية كبيرة عند إنشاء جداول الأرقام القياسية.

من عيوب الوسط الهندسي:

1- لا يمكن حسابه للجداول المفتوحة

2- لا يمكن حسابه إذا كانت إحدى القيم سالبة أو صفراً.

3- هو قيمة مجردة وغير واضحة في بعض الحالات .

4- صعوبة حسابه وخاصة إذا كانت البيانات كبيرة جداً.

7.2 العلاقة بين الوسط الحسابي والوسيط والمنوال:

1- **الوسط الحسابي:** المسافة تحت المنحنى تضم مجموعة علاقات وتكون موزعة بالتساوي على طرفيه أي أن مجموع الانحرافات السالبة على الجانب الأيسر مساوية لمجموع الانحرافات الموجبة على الجانب الأيمن فهو يمر من النقطة المركزية للمسافة تحت المنحنى .

2- **الوسيط:** يقسم المساحة تحت المنحنى إلى قسمين متساويين بحيث أن البيانات التي تزيد على قيمة الوسيط تساوي عدد البيانات التي تقل عن قيمة الوسيط.

3- **المنوال**: قيمته تطابق أعلى نقطة على المنحنى.

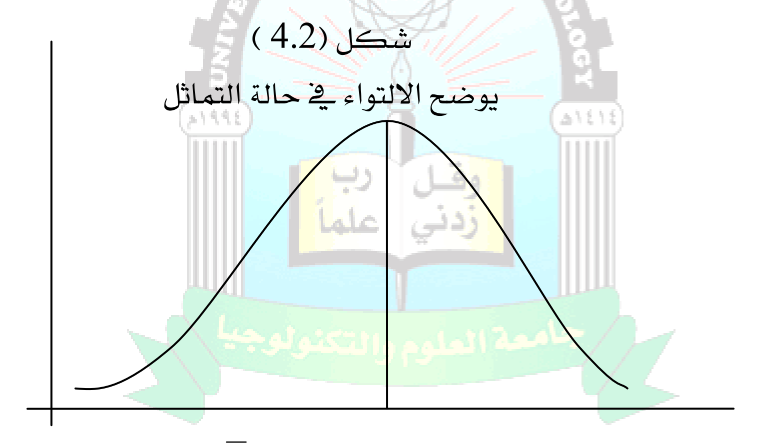
4- تتطابق قيمة الوسط الحسابي والوسيط والمنوال فقط عندما يكون التوزيع التكراري متماثل أي أن: الوسط الحسابي = الوسيط = المنوال

في التوزيعات المتماثلة التي لها أكثر من منوال فإن الوسط الحسابي أو الوسيط يكون هو المقياس المناسب.

عندما يكون التوزيع شديد الالتواء فإن الوسيط غالباً ما يكون أفضل مقاييس الوسط لأنه يقع بين الوسط الحسابي والمنوال.

شكل (4.2)

يوضح الالتواء في حالة التماثل



$$\overline{X} = M_d = M_o$$

وفي حالة التوزيعات الملتوية التواءً بسيطاً أي قريبة من التماثل فإن الوسط الحسابي – المنوال = 3(الوسط الحسابي – الوسيط) .

ومن الشكلين(5.2) و(6.2) نلاحظ أن:

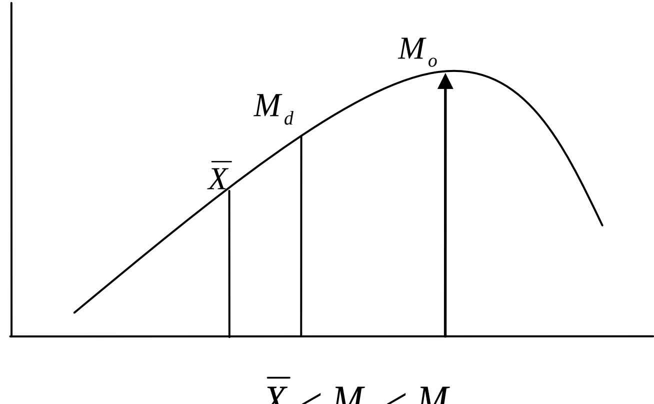
1- في حالة التوزيع غير المتماثل السالب الالتواء يكون المنوال أكبر من كل من الوسيط والوسط الحسابي.

2. حالة عدم التماثل (التواء سالب)

في التوزيعات التي تحتوي قيماً شاذه أو متطرفة فإن الوسيط أو المنوال هما المقياسان المناسبان وإذا كان حجم البيانات كبيرة فإن الوسط الحسابي هو الأسهل لكنه ليس الأفضل.

شكل (5.2)

يوضح الالتواء السالب

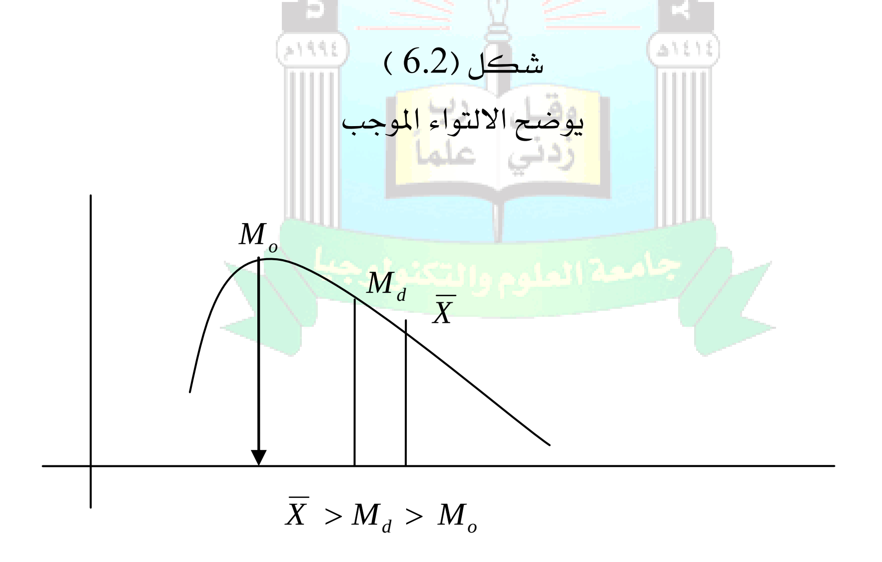


$$\overline{X} < M_d < M_o$$

2- في حالة التوزيع غير المتماثل الموجب الالتواء يكون المنوال أصغر من كل من الوسيط والوسط الحسابي.

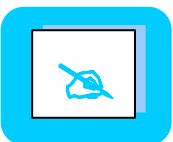
شكل (6.2)

يوضح الالتواء الموجب



$$\overline{X} > M_d > M_o$$

مثال(21.2) :



إذا كان متوسط التوزيع 1.63 وكان وسيطه 1.61 احسب المنوال ثم حدد نوع الالتواء

الحل:

نعلم أن : الوسط الحسابي – المنوال = 3 (الوسط الحسابي – الوسيط)

المنوال = الوسط – 3 (الوسط – الوسيط)

المنوال = 1.63 – 3(1.63 – 1.61)

المنوال = 1.63 – 3 (0.02)

المنوال = 1.63 – 0.06 = 1.57

وحيث أن المنوال أصغر من الوسيط وأصغر من الوسط الحسابي فيكون الالتواء موجب باتجاه اليمين.



تدريب (8)



احسب المتوسط الحسابي لجدول تكراري إذا علمت أن:

المنوال = 156.647 والوسيط = 155.536

3. الربيعات والعشيريات والمئينات
Deciles, Quartiles and Percentiles

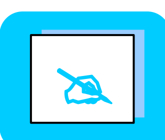
إذا كانت مقاييس النزعة المركزية تستهدف قياس أو تحديد المركز الذي تتمحور حوله البيانات وأن جميعها تتمثل بقيمة مفردة واحدة تكون ممثلة لبيانات العينة. فإن هناك مقاييس أخرى تشبه الوسيط في أنها مقاييس للوضع وفي طريقة حسابها ، ولكنها ليست من المتوسطات. وهذه المقاييس هي الربيعيات و التي تقسم المفردات إلى أربعة أجزاء كل جزء منها يشتمل على عدد متساوي من المفردات ويرمز له بالرمز $Q_i\ (i=1,2,3)$ ويلاحظ أننا لا نحتاج لحساب الربيع الرابع لأنها ستكون مكملة للربيعات $Q_1\ ,\ Q_2\ ,Q_3$. وكذلك يمكن تقسيم البيانات إلى عشرة أقسام متساوية تسمى بالعشيريات ويرمز لها بالرمز $D_i\ \ (i=1,2,......,9)$ ، كما يمكن تقسيم البيانات إلى مائة قسم متساوٍ تسمى بالمئينات ويرمز لها بالرمز $P_i\ \ (i=1,2\ ,3,.......,99)$ ولحساب هذه المقاييس في حالة البيانات الخام (غير المبوبة) **عزيزي الدارس** نقوم بتحديد موقع الربيع أو العشير أو المئين من الصيغ الآتية:

موقع الربيع $\dfrac{(i)(n+1)}{4}\ ,\ i=1,2,3$

موقع العشير $\dfrac{(i)(n+1)}{10}\ ,\ i=1,2,3\ ,.....,9$

موقع المئين $\dfrac{(i)(n+1)}{100}\ ,\ i=1,2,3,.......,99$

مثال (22.2):



للبيانات التالية $6\ ,\ 18\ ,\ 15,\ 9\ ,\ 11\ ,\ 8\ ,\ 14$ احسب

$Q_1\ ,\ Q_3\ ,\ M_d\ ,D_5\ ,\ P_{30}$

الحل:

نرتب البيانات تصاعدياً $6\ ,\ 8\ ,\ 9\ ,\ 11\ ,\ 14\ ,\ 15\ ,\ 18$

أولاً: حساب Q_1

حيث أن عدد المفردات فردياً فإننا نقوم بالآتي:

تحديد موقع الربيع الأول من العلاقة: $\dfrac{(i)(n+1)}{4} = \dfrac{(1)(7+1)}{4} = 2$

إذاً العنصر الثاني يمثل الربيع الأول أي $Q_1 = 8$

ثانياً: حساب Q_3

نحدد موقع الربيع الثالث من العلاقة: $\frac{(i)(n+1)}{4} = \frac{(3)(7+1)}{4} = 6$

فيكون العنصر الذي ترتيبه 6 يمثل الربيع الثالث أي أن $Q_3 = 15$

ثالثاً: حساب الوسيط M_d وهو القيمة التي تتوسط التوزيع وهي العنصر الرابع وبذلك تكون قيمة الوسيط $M_d = 11$

رابعاً: حساب D_5

نحدد موقع العشير الخامس من العلاقة : $\frac{(i)(n+1)}{10} = \frac{(5)(7+1)}{10} = 4$

فيكون العنصر الرابع يمثل العشير الخامس أي أن $D_5 = 11$ وهي نفس قيمة الوسيط.

خامساً: حساب قيمة P_{30}

نحدد موقع المئين الثلاثين من العلاقة $\frac{(i)(n+1)}{100} = \frac{(30)(7+1)}{100} = 2.4$

وبعد الحصول على الموقع 2.4 نجد أن قيمة المئين الثلاثين هي العنصر الثاني مضافاً إليه حاصل ضرب 0.4 في حاصل الفرق بين القيمة الثالثة والثانية أي أن

$P_{30} = 8 + (9-8) * .4 = 8.4$

أما في حالة التوزيعات التكرارية فإننا سوف نحدد موقع كل مقياس بنفس الطريقة للبيانات غير المبوبة إلا أننا سوف نستخدم $\sum f_i$ بدلاً من n وتحسب الربيعات والعشيريات والمئينات من العلاقات الآتية:

1- الربيعات $Q_i = L + \frac{\frac{(i)\sum f_i}{4} - f_1}{f_2} \cdot H \ , \quad i = 1,2,3$

2- العشيريات $D_i = L + \frac{\frac{(i)\sum f_i}{10} - f_1}{f_2} \cdot H \ , \quad i = 1,2,3,\dots\dots,9$

3- المئينات $P_i = L + \frac{\frac{(i)\sum f_i}{100} - f_1}{f_2} \cdot H \ , \quad i = 1,2,3,\dots\dots,99$

حيث:

L : الحد الأدنى للفئة الربيعية أو العشيرية أو المئينية .

f_1 : لتكرار المتجمع الصاعد السابق لموقع المقياس

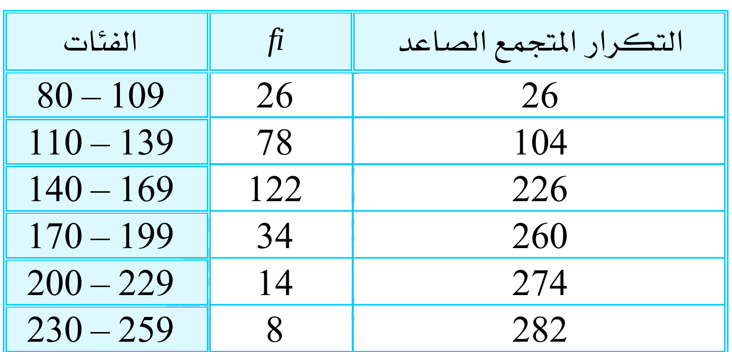
f_2 : تكرار الفئة الربيعية أو العشيرية أو المئينية.

H : طول الفئة الربيعية أو العشيرية أو المئينية

مثال(23.2):

من بيانات الجدول الآتي :

جدول (5.2)



الفئات	fi	التكرار المتجمع الصاعد
80 – 109	26	26
110 – 139	78	104
140 – 169	122	226
170 – 199	34	260
200 – 229	14	274
230 – 259	8	282

* المئين الخامس والعشرين يدعى بالربيع الأول
* المئين الخمسين يـــدعى بالربيع الثاني أو الوسيط
* المئين الخامس والسبعين يدعى بالربيع الثالث

أوجد:

1. العشير الأول والثاني D_1 , D_2
2. الربيع الأول والثالث Q_1 , Q_3
3. المئين P_{30}

الحل:

1. نحدد موقع D_1 , D_2

موقع العشير الأول $\frac{1(282)}{10} = 28.2$

موقع العشير الثاني $\frac{2\,(282)}{10} = 56.4$

فتكون قيمة العشير الأول $D_1 = 110 + \frac{28.2 - 26}{78}.(30) = 110.846$

و قيمة العشير الثاني $D_2 = 110 + \frac{56.4 - 26}{78}.(30) = 121.69$

2- نحدد موقع الربيع الأول $\frac{1(282)}{4} = 70.5$

$$\frac{3(282)}{4} = 211.5$$ موقع الربيع الثالث

فتكون قيمة الربيع الأول والثالث هي:

$$Q_1 = 110 + \frac{70.5 - 26}{78}(30) = 127.115$$

$$Q_3 = 140 + \frac{211.5 - 104}{122}(30) = 166.434$$

المدى الربيعي : هو الفرق بين الربيع الثالث والربيع الأول أي أن:
المدى الربيعي = المئين **75**− المئين **25**.

3. قيمة المئين الثلاثين هي:

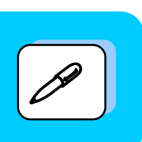
نحدد موقع المئين الثلاثين

$$\frac{30\ (282)}{100} = 84.6$$

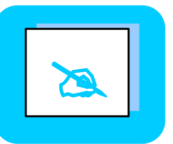
$$p_{30} = L + \frac{\frac{30\sum f_i}{100} - f_1}{f_2} H$$

$$p_{30} = 110 + \frac{84.6 - 26}{78}.30 = 132.54$$

تدريب (9)



وضح كيف يمكن تحديد موقع الربيع أو العشير أو المئين، وهل هي تعتبر من مقاييس النزعة المركزية في الجدول الآتي:



تمرين عام (محلول):

تم سحب عينة عشوائية بحجم 200 طالب من طلاب المستوى الثاني فتبين أن 80 منهم باقون و 120 مستجدون كما تم تبويب درجاتهم في مادة معمارية الحاسوب كما في الجدول التالي :

فئات درجات الطلاب	عدد الطلاب		إجمالي
	مستجدون	باقون	
0 – 10	5	2	7
10 - 20	6	5	11
20 - 30	7	10	17
30 - 40	8	15	23
40 - 50	14	17	31
50 - 60	20	20	40
60 - 70	30	4	34
70 - 80	15	3	18
80 - 90	10	2	12
90 - 100	5	2	7
$\sum$	120	80	200

أوجد :

1- متوسط درجات الطلاب الباقون والمستجدون.

2-متوسط درجات الطلاب ككل في العينة.

3- احسب الوسيط للطلاب ككل.

4- احسب المنوال

5- حدد طبيعة التوزيع لعينة الطلاب ككل

زدني علما
جامعة العلوم والتكنولوجيا

الحل:

جدول (6.2)

الفئات	عدد الطلاب		مراكز الفئات	$f_i\ x_i$	$f_i\ x_i$
	باقون f_i	مستجدون f_i	x_i	باقون	مستجدون
0 – 10	2	5	5	10	25
10 - 20	5	6	15	75	90
20 - 30	10	7	25	250	175
30 - 40	15	8	35	525	280
40 - 50	17	14	45	765	630
50 - 60	20	20	55	1100	1100
60 - 70	4	30	65	260	1950
70 - 80	3	15	75	225	1125
80 - 90	2	10	85	170	850
90 - 100	2	5	95	190	495
$\sum$	80	120		3570	6700

1. المتوسط الحسابي للطلاب الباقون.

$$\bar{X} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i} = \frac{3570}{80} = 44.625$$

1. المتوسط الحسابي للطلاب المستجدون

$$\bar{X} = \frac{\sum f_i\ x_i}{\sum f_i} = \frac{6700}{120} = 55.833$$

ولحساب الوسط الحسابي للطلاب ككل نقوم بجمع الطلاب الباقين والمستجدين في مجموعة واحدة ونكون جدولاً جديداً يضم البيانات لجميع الطلاب كما هو موضح في الجدول(7.2)

جدول(7.2)

الفئات	f_i	x_i	$f_i\ x_i$	تكرار متجمع صاعد
0 – 10	7	5	35	7
10 - 20	11	15	165	18
20 - 30	17	25	425	35
30 - 40	23	35	805	58
40 - 50	31	45	1395	89
50 - 60	40	55	2200	129
60 - 70	34	65	2210	163
70 - 80	18	75	1350	181
80 - 90	12	85	1020	193
90 - 100	7	95	665	200
$\sum$	200		10270	

2. الوسط الحسابي بعد ضم الطلاب الباقون والمستجدون في مجموعة واحدة من العلاقة:

$$\overline{X} = \frac{\sum f_i\ x_i}{\sum f_i} = \frac{10270}{200} = 51.35$$

3. لحساب الوسيط للطلاب ككل نتبع الآتي:

أ. نكون جدول تكرار متجمع صاعد كما في الجدول (7.2)

ب. نحدد موقع الوسيط من العلاقة : $\frac{\sum f_i}{2} = \frac{200}{2} = 100$

ج. نحدد الفئة الوسيطية هي (60 – 50) فيكون الحد الأدنى للفئة الوسيطة

$L = 50$

وكذلك نحدد قيمة $f_1 = 89$ وهو التكرار المتجمع الصاعد السابق لموقع الوسيط وقيمة $f_2 = 40$ وهو تكرار الفئة الوسيطية، وطول الفئة $H = 10$

د. نطبق العلاقة الخاصة بحساب الوسيط من بيانات مبوبة

$$\because M_d = L + \frac{\frac{\sum f_i}{2} - f_1}{f_2} * H$$

$$\therefore M_d = 50 + (\frac{100-89}{40}) * 10 = 50 + \frac{11}{40} * 10$$

$$M_d = 50 + \frac{110}{40} = 50 + 2.75 = 52.75$$

4. يتم حساب المنوال من خلال الصيغة التالية:

$$M_0 = L + (\frac{d_1}{d_1 + d_2}) * H$$

حيث نلاحظ أن الفئة المنوالية هي (50 – 60) لأنه يقابلها أكبر تكرار (40)فتكون قيمة $L = 50$ ومن الجدول(7.2) أن $H = 10$, $d_2 = 6$, $d_1 = 9$

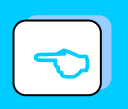
$$\therefore M_0 = 50 + (\frac{9}{9+6}) * 10 = 50 + \frac{90}{15}$$

$$M_o = 50 + 6 = 56$$

5. من النتائج السابقة في 3و4 نلاحظ أن الوسط الحسابي أقل من كل من الوسيط والمنوال أي $\left[\begin{array}{l} \overline{X} < M_d < M_o \\ 51.35 < 52.75 < 56 \end{array}\right]$

وعليه فإن التوزيع ملتوي باتجاه اليسار التواء سالب.

نشاط



قم بزيارة للجهاز المركزي للإحصاء وتعرف على عملية جمع البيانات الخاصة بالسكان في الجانب التعليمي لكي تتعرف على كيفية تحديد نسبة التعلم بين الإناث وكذلك بين الذكور. محاولاً عمل مقارنة بين محافظات الجمهورية اليمنية في هذا المجال.

أسئلة التقويم الذاتي (3) :

- أكمل الفراغات الآتية بالكلمات المناسبة:

1- إذا كان التوزيع ملتوي فإن أفضل مقاييس المتوسطات هو.......................

2- يعرف الوسيط بأنه ....

3- يتأثر الوسيط بالقيم .... بينما يتأثر الوسط الحسابي بالقيم..................

4- تعرف الفئة المنوالية بأنها...

5- مجموع انحرافات القيم عن وسطها الحسابي ....

2- أي مقاييس النزعة المركزية أكثر استعمالاً ؟ ولماذا ؟

3- احسب الوسط الحسابي والوسيط والمنوال للبيانات التالية:

1) 5 , 2 , 0 , - 3 , -1

2) 0.004, -.002 , 0.003 , 0.001

3) 2 , 2 , 3 , 7 , 10 , 100

4- متوسط 23 مفردة هو 14.7 فما هو مجموع هذه القيم ؟

5- فيما يلي مستوى السكر في الدم مأخوذاً في الصباح قبل تناول الفطور لعشرة أطفال 56 , 62 , 63 , 65 , 65 , 65 , 68 , 70 , 72, 66

احسب الوسط الحسابي والوسيط والمنوال لهذه القيم.

6- جدول التوزيع التكراري التالي يلخص مجموعة بيانات عن درجة تلوث الهواء(مقاسه بالكيلوجرام في المتر المكعب) في 53 منطقة سكنية

الفئات	10-	20 -	30 -	40 -	50-	60-	70-80
التكرارات	5	17	9	12	4	4	2

أحسب الوسط الحسابي والوسيط والمنوال.

؟

4. مقاييس التشتت Measures of Dispersion

عزيزي الدارس، لقد ناقشنا سوياً في القسم السابق من هذه الوحدة مقاييس النزعة المركزية وكيفية قياس صفة التركيز في البيانات وتحديد الاتجاهات العامة لها وتبسيطها عن طريق تلخيصها في قيمة واحدة بهدف قياس أو تحديد المركز الذي تتمحور حوله المفردات والتي من خلالها توصف باقي المفردات أو المعلومات من خلال قياس المتوسطات (مقاييس النزعة المركزية) إلا أنها ليست كافية لتشكيل صورة ذهنية متكاملة عن التوزيع التكراري لوصف وتلخيص مجموعة من البيانات بشكل كامل لأن الوصف يجب أن يكون من خلال أكثر من صفة واحدة فمقاييس النزعة المركزية لا تعطي وصفاً لدرجة التركيز كما أنها لا تعطي وصفاً لاتجاه هذا التركيز هل هو نحو القيم الصغيرة أم نحو القيم الكبيرة كما أنها لا تبين إلى أي مدى تتركز القيم حول وسطها الحسابي.

ووصف البيانات بشكل كامل يحتم علينا دراسة أكثر من جانب عنها ، فما زالت الحاجة ماسة لمقاييس أو أخرى تحدد مدى تقارب أو تباعد مفردات الظاهرة عن بعضها البعض حيث تقاس قيم التركز بمقاييس النزعة المركزية وتقاس درجة التجانس أو التركز بمقاييس التشتت وتقاس اتجاهات التركز وإبعاده بواسطة مقاييس الالتواء أو التفلطح . والإحصائي يهتم بدرجة التركز أو التجانس بنفس درجة اهتمامه بمعرفة قيم التركز لأن التشتت والتركز من أهم معالم وصفات المجتمع الإحصائي.

من مقاييس التشتت:
- المدى.
- الانحراف الربيعي.
- الانحراف المتوسط.
- التباين.
- الانحراف المعياري.

أما مقاييس التشتت النسبي:
- معامل الاختلاف.
- القيم المعيارية.

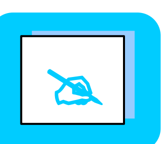
1.4 مفهوم التشتت:

يقصد بالتشتت من الناحية الإحصائية بأنه مقدار التقارب والتغيير بين المفردات أي مدى تقارب أو تباعد قيم مفردات رقمية لظاهرة ما حول أي قيمة متوسطة لتلك البيانات (المفردات).وهذه القيمة قد تكون الوسط الحسابي أو الوسيط أو المنوال.

وتنبع أهمية قياس التشتت من حقيقة أنه ربما تتساوى المتوسطات لأكثر من مجموعة إلا أن هذه المجموعات تكون مختلفة كثيراً من حيث تجانس البيانات ولذا من الخطأ أن نقول أن هذه المجموعات متساوية والمثال الآتي يوضح ذلك.

جامعة العلوم والتكنولوجيا

مثال:



إذا كان لدينا ثلاث ورش (A, B , C) وقمنا بسحب عينة عشوائية من أربعة عمال من كل ورشة وسجلنا أجور العمال في اليوم في تلك الورش فكانت كالآتي :

	الورشةA	الورشة B	الورشة C
العامل الأول	200 ريال	350 ريال	350 ريال
العامل الثاني	400 ريال	450 ريال	350 ريال
العامل الثالث	300 ريال	300 ريال	350 ريال
العامل الرابع	500 ريال	300 ريال	350 ريال

وعند حساب متوسط الأجر للعامل الواحد نجده يساوي 350 ريال في جميع الورش ولو اعتمدنا على الوسط الحسابي فقط لوصف هذه البيانات لوصلنا إلى نتيجة مضلله وهي أن أجور العمال في هذه الورش تتوزع توزيعاً متشابهاً ولكن الواقع غير ذلك لأن البيانات في الورشة C متجانسة تمام التجانس أي أن درجة التشتت فيها تساوي الصفر وبالتالي فإن وسطها الحسابي يمثلها تمثيلا كاملاً.

أما في الورشة B فالأمر يختلف كثيراً حيث تتقارب البيانات أكثر من الورشة A ولذا فإن درجة تشتت أجور تلك الورشة ضعيفة أي أضعف مما هي عليه في الورشة A .

والتشتت لأي مجموعة من البيانات يقصد به التباعد أو الاختلاف بين مفرداتها ، فتكون درجة التشتت إما معدومة أي صفر (إذا تساوت جميع القيم) أو ضعيفة (قريبة من الصفر) أو كبيرة (بعيدة كثيراً عن الصفر) ويجب ألا تكون قيمة التشتت أقل من الصفر حيث لا معنى للتشتت السالب. ونخلص من ذلك إلى أهمية كل من مقاييس النزعة المركزية ومقاييس التشتت جنباً إلى جنب عند تحليل البيانات الرقمية وسوف نستعرض فيما يلي مقاييس التشتت الآتية:

2.4 المدى The Range

تعريف(1): المدى هو الفرق بين أكبر قيمة وأصغر قيمة في البيانات، وكلما كان هذا الفرق صغيراً كلما كان دليلاً على مدى تجانس البيانات أي انخفاض التشتت.

تعريف(2): المدى في حالة البيانات المبوبة (التوزيعات التكرارية) هو: الفرق بين الحد الأعلى لآخر فئة مقفلة والحد الأدنى لأول فئة مقفلة. أو هو الفرق بين مركز الفئة الأولى ومركز الفئة الأخيرة في التوزيعات المقفلة.

ومن هذين التعريفين يتبين أن المدى لا يعتمد على جميع البيانات بل يعتمد على أكبر وأصغر قيمة في البيانات فقط. وهذا يقلل من أهميته. لأنه قد تكون القيمتان(أكبر وأصغر قيمة) شاذتين، عندئذ يكون المدى كبيراً بينما تكون مفردات البيانات أصلاً ليست متباعدة عن بعضها البعض فمثلاً إذا كانت أعمار مجموعة من الموظفين في إحدى المؤسسات هي:

$$23 , 37 , 38 , 41 , 36 , 40 , 65$$

فإن المدى هو $42 = 23 - 65$ ، بينما تلاحظ أن معظم الأعمار واقعة بين 36 و 41 سنة أي أنها متقاربة من بعضها البعض. وهذا يشير إلى أن المدى عنصر لا يُعتمد عليه كثيراً كمقياس لتشتت البيانات. وللتغلب على نقطة ضعف المدى وهي تأثره بالقيم المتطرفة، يتم حذف القيمتين المتطرفتين من المجموعة وهما 65 و 23 فيصبح المدى $5 = 36 - 41$ لأن اعتماده عليها يجعله حساساً لحجم العينة، فعادة ما تزداد قيمة المدى بزيادة حجم العينة.

ويستخدم المدى في خرائط مراقبة جودة الإنتاج حيث يتطلب الأمر معرفة درجة التشتت لعدد كبير من العينات المسحوبة عشوائياً من مجتمع الوحدات الإنتاجية، ويمكن تجنب عيوب المدى باستخدام الانحراف المعياري، كما يمكن معالجة حساسية المدى للقيم المتطرفة بحساب المدى بين المئين التسعين والمئين العاشر أو بحساب المدى الربيعي.

ومن عيوب المدى:

1- أنه يتأثر بالقيم المتطرفة لأنه يعتمد في حسابه على أكبر وأصغر قيمة فهو مقياساً مضللاً.

2- لا يمكن حسابه للجداول التكرارية المفتوحة .



تدريب (10)

أحسب المدى للقيم التالية:

22،20،19،17،12،10،13،25،15

1.2.4 المدى الربيعي Interlunation Range

هو الفرق بين الربيع الأعلى والربيع الأدنى أي أن:

المدى الربيعي = الربيع الثالث – الربيع الأول

$= Q_3 - Q_1$

1.1.2.4 الانحراف الربيعي (نصف المدى الربيعي):

يمكن التقليل من أثر تأثير المدى المطلق بالقيم المتطرفة والشاذة بإيجاد نصف المدى الربيعي الذي يعرف بأنه القيمة التي تنحرف بها نقاط الربيع الأول (الأدنى) والربيع الثالث(الأعلى) عن الوسيط، ويعرف باسم نصف المدى الربيعي ويحسب من

$\frac{Q_3 - Q_1}{2}$ = نصف المدى الربيعي.

خطوات حساب نصف المدى لربيعي:

1- نحدد رتبتي الربيع الأول والثالث فيكون:

ترتيب الربيع الأول هو $\frac{n}{4}$ إذا كان عدد المفردات (n) وترتيب الربيع الثالث هو $\frac{3n}{4}$

2- نحسب قيمة الربيعين الأول والثالث بطريقة حساب الوسيط من التوزيعات التكرارية نفسها.

باستخدام المعادلة التالية:

$$Q_i = L + \left(\frac{\frac{(i)\sum f_i}{4} - f_1}{f_2}\right) * H$$

مثال (24.2):

احسب الانحراف الربيعي للتوزيع التكراري التالي:

فئات	10-	15-	20-	25-	30-	35-	40-	45-	50– 55
تكرارات	4	6	12	20	18	14	14	2	10

الحل:

$$\text{ترتيب الربيع الأول} = \frac{100}{4} = 25$$

$$\text{ترتيب الربيع الثالث} = \frac{100\times3}{4} = 75$$

جدول (8.2)

الفئات	التكرار f_i	التكرار المتجمع الصاعد
10-	4	4
15-	6	10
20-	12	22
25-	20	42
30-	18	60
35-	14	74
40-	14	88
45-	2	90
50 - 55	10	100

قيمة الربيع الأول $Q_1 = 25 + (\frac{25-22}{20}) * 5 = 25.75$

قيمة الربيع الثالث $Q_3 = 40 + (\frac{75-74}{14}) * 5 = 40.36$

نصف المدى الربيعي= $\frac{25.75 - 40.36}{2} = 7.31$

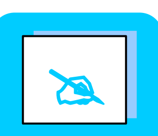
والسؤال الذي يطرح نفسه: لماذا لا نبحث عن مقياس مناسب للتشتت تسهم في تشكيله كل قيمة من القيم الإحصائية بدلاً من أن يقتصر على أكبر قيمة وأصغر قيمة أو على المئين ؟ ومن الواضح أن التشتت يعود إلى قرب أو بعد القيم عن وسطها الحسابي فلنحاول إذاً التعبير عن التشتت بدلالة انحراف القيم عن متوسطها الحسابي أي بدلالة $d_i = x_i - \bar{x}$. ونعلم من خواص المتوسط أن مجموع هذه الانحرافات عن الوسط يساوي الصفر مما لا يترك مجالاً للتفكير في متوسط هذه الانحرافات كمقياس للتشتت . وحل هذه المشكلة سهل طالما أنه يعود إلى وجود انحرافات موجبة وأخرى سالبة فلماذا لا نحسب متوسط القيم المطلق للانحرافات؟

3.4 الانحراف المتوسط (The Mean Deviation):

هو عبارة عن معدل الانحرافات المطلقة لكافة البيانات X_i عن قيمة الوسط الحسابي $\bar{X}$ لتلك البيانات مع إهمال الإشارة الجبرية. وهو أحد مقاييس التشتت الهامة والتي تستخدم في التعرف على مدى تجانس مجموعة من المفردات لأنه كلما كانت قيم المفردات متجانسة كانت الفروق بينها صغيرة وكانت انحرافاتها عن متوسطها الحسابي صغيرة ويحسب من العلاقة:

$$\text{M. D} = \frac{\sum_{i=1}^{n} |X_i - \bar{X}|}{n}$$

حيث $\bar{x}$ تمثل الوسط الحسابي ، X_i تمثل البيانات.



مثال(25.2):

احسب الانحراف المتوسط للنفقات الشهرية للأسر بآلاف الريالات

7،12،5،6،4،3،8،3

الحل : فيما يلي طريقة حساب الانحراف المتوسط .

i	X_i	$\lvert X_i - \bar{X} \rvert$
1	12	6
2	8	2
3	7	1
4	6	0
5	5	1
6	4	2

جامعة العلوم والتكنولوجيا

7	3	3
8	3	3
$\sum$	48	18

$= 6\bar{X}$

$$M.D = \frac{\sum |X_i - \bar{X}|}{n} = \frac{18}{8} = 2.25$$

مثال(26.2):



احسب الانحراف المتوسط لقيم المجموعة الآتية:

9 , 6 , 8 , 10 , 12

i	X_i	$\lvert X_i - \bar{X} \rvert$
1	9	0
2	6	3
3	8	1
4	10	1
5	12	3
$\sum$	45	8

الحل:

$$\bar{X} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{45}{5} = 9$$

$$M.D = \frac{\sum |X_i - \bar{X}|}{n} = \frac{8}{5} = 1.6$$

حساب الانحراف المتوسط في حالة التوزيع التكراري (بيانات مبوبة) .

يمكن حساب الانحراف المتوسط من التوزيع التكراري بإتباع الخطوات التالية:

1- حساب الوسط الحسابي كما بينا سابقاً.

2-حساب الفروق المطلقة بين مراكز الفئات والوسط الحسابي أي $|X_i - \bar{X}|$.

4- ضرب الناتج من الخطوة السابقة في التكرارات $f_i |X_i - \bar{X}|$.

5- جمع العمود $f_i |X_i - \bar{X}|$

6- نقسم الناتج على مجموع التكرارات فيكون الناتج هو الانحراف عن المتوسط.

مثال(27.2):

أوجد الانحراف المتوسط لبيانات الجدول التالي والذي يوضح الدخل الشهري لـ 100 أسرة

الفئات	5-	10-	15-	20-	25-	30-	35 - 40
التكرارات	10	10	20	30	20	5	5

الحل

جدول (9.2)

الفئات	f_i	X_i	$X_i f_i$	$\lvert X_i - \bar{X}\rvert$	$f_i\lvert X_i - \bar{X}\rvert$
5-	10	7.5	75	13.75	137.5
10-	10	12.5	125	8.75	87.5
15-	20	17.5	350	3.75	75
20-	30	22.5	675	1.25	37.5
25-	20	27.5	550	6.25	125
30-	5	32,5	162.5	11.26	56.25
35 - 40	5	37,5	187.5	16.25	81.25
المجموع	100		2125		600

$$\bar{X} = \frac{\sum_{i=1}^{k} f_i X_i}{\sum f_i} = \frac{2125}{100} = 21.25$$

$$M.D = \frac{\sum_{i=1}^{k} f_i\lvert X_i - \bar{X}\rvert}{\sum_{i=1}^{k} f_i}$$

$$M.D = \frac{600}{100} = 6$$

3.3.4 من خواص الانحراف المتوسط:

1- إهمال الإشارة الجبرية يجعل من الصعب التعامل معه رياضياً

2- يعتبر من أحسن المقاييس لتمثيل التشتت

جامعة العلوم والتكنولوجيا

3- سهل جداً للفهم

4- مفهوم وسطي يأخذ قيم كافة المفردات بعين الاعتبار عند حسابه.



تدريب (11)

1. علل: من الصعب التعامل مع الانحراف المتوسط رياضياً.
2. احسب الانحراف المتوسط لقيم المجموعة الآتية :

35 , 20 , 30 , 40 50

4.4 التباين والانحراف المعياري (Variance and Standard Deviation)

1.4.4 التباين (The variance) :

إذا كانت نقطة الضعف الأساسية في الانحراف المتوسط هي أنه يهمل الإشارة الجبرية، ويأخذ كل الانحرافات على أنها موجبة، فهل بالإمكان التغلب على ذلك العيب، ولعل من أقرب الحلول لمعالجة تلك المشكلة هي أن نقوم بتربيع انحرافات القيم عن وسطها الحسابي وبذلك نتغلب على هذا العيب، وإذا قمنا بذلك فإننا نكون قد وصلنا إلى مقياس آخر من مقاييس التشتت وهو ما يعرف بالتباين.

تعريف التباين هو: متوسط مجموع مربعات انحرافات القيم عن وسطها الحسابي ويرمز له بالرمز S^2 في حالة حسابه للعينات، أما في حالة حسابه للمجتمع الإحصائي فنرمز له بالرمز σ^2 (ينطق سيجما تربيع). ويحسب من العلاقة:

$$S^2 = \frac{(x_1-\bar{x})^2 + (x_2-\bar{x})^2 + \dots + (x_n-\bar{x})^2}{n} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n}$$

أما في حالة المعاينة فيتم طرح (1) من n للتعويض عن خطأ المعاينة وتصبح المعادلة بالصورة:

$$S^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}$$



مثال(28.2):

احسب التباين للبيانات التالية 2، 3، 4،7

الحل: 1- نحسب الوسط الحسابي $\bar{X} = \frac{16}{4} = 4$

2- نحسب انحرافات القيم عن وسطها الحسابي كما يُبينها الجدول الآتي:

i	X_i	$X_i - \bar{X}$	$(X_i - \bar{X})^2$
1	2	-2	4
2	3	-1	1
3	4	0	0
4	7	3	9
$\sum$	-	0	14

3- نطبق العلاقة:

$$S^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n} = \frac{14}{4} = 3.5$$

ونقطة الضعف الأساسية في التباين هي اعتماده على مربعات انحرافات القيم عن وسطها الحسابي في قياس درجة التشتت أي أنه يعتمد على قيم غير أصلية ويمكن تلافي هذا الضعف بأخذ الجذر التربيعي للتباين فنحصل على ما يسمى بالانحراف المعياري.

2.4.4 **الانحراف المعياري** (Standard Deviation):

هو أحد مقاييس التشتت الهامة وأكثرها شيوعاً و استخداماً في علم الإحصاء ويرمز له بالرمز S إذا كانت البيانات تمثل عينة من مجتمع و σ إذا كانت البيانات تمثل المجتمع الإحصائي، و يعرف بأنه الجذر التربيعي للتباين .

1.2.4.4 **طرق حساب الانحراف المعياري** :

أ- في حالة البيانات غير المبوبة: يحسب من الصيغة:

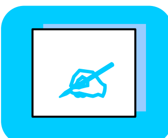
$$S = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n}}$$

أما في حالة البيانات المبوبة (التوزيع التكرارية) فيُحسب الانحراف المعياري من العلاقة التالية:

$$S=\sqrt{\frac{\sum_{i=1}^{k} f_i X_i^2}{\sum_{i=1}^{k} f_i}-\left(\frac{\sum_{i=1}^{k} f_i X_i}{\sum f_i}\right)^2}$$

حيث x_i تمثل مراكز الفئات.

مثال(29.2):



الجدول التالي يتضمن نتائج اختيار 33 شخص من المتقدمين للحصول على عمل في إحدى الشركات والمطلوب إيجاد التشتت للدرجات باستخدام طريقة الانحراف المعياري .

فئات الدرجات	22 – 32	33 – 43	44 – 54	55 – 65	66 – 76
التكرارات f_i	4	6	10	9	4

الحل :

نوجد مراكز الفئات X_i ومربعاتها حسب الجدول(2-9)

جدول (9.2)

الفئات	f_i	X_i	$X_i f_i$	X_i^2	$f_i X_i^2$
22 – 32	4	27	108	729	2916
33 – 43	6	38	228	1444	8664
44 – 54	10	49	490	2401	24010
55 – 65	9	60	540	3600	32400
66 – 76	4	71	284	5041	20164
$\sum$	33	--	1650	--	88154

نطبق العلاقة :

$$S=\sqrt{\frac{\sum_{i=1}^{k} f_i X_i^2}{\sum_{i=1}^{k} f_i}-\left(\frac{\sum_{i=1}^{k} f_i X_i}{\sum_{i=1}^{k} f_i}\right)^2}$$

$$=\sqrt{\frac{88154}{33}-\left(\frac{1650}{33}\right)^2}$$

$$=\sqrt{2671.33-2500}$$

$$=\sqrt{171.33}$$

لتصحيح (شبرد) معامل 2.2.4.4 $=13.089$

التباين:

عند حساب الانحراف المعياري S فإنه يكون معرضاً لبعض الأخطاء الناتجة من تجميع البيانات في فئات وتسمى بأخطاء التجميع.

ولتعديل هذا الخطأ يمكن استخدام ما يدعى بمعامل (شبرد) وهو

التباين المعدل = التباين من المفردات المجمعة - $\frac{C^2}{12}$

حيث

C : طول الفئة

$\frac{C^2}{12}$: معامل التصحيح

ولحساب التباين المعدل S'^2 للمثال السابق نلاحظ أن:

$$S'^2 = (13.089)^2 - \frac{(11)^2}{12}$$

$$= 171.3219 - 10.08333$$

$$= 161.23857$$

الانحراف المعياري المصحح أو المعدل هو: $S' = 12.698$

3.2.4.4 بعض خصائص الانحراف المعياري

1- أنه مقياس محدد جبرياً ويتمتع ببعض المزايا الجبرية

2- لا يمكن حسابه للجداول التكرارية المفتوحة

3- إذا أضفنا أو طرحنا مقدار ثابت C إلى كل قيم X_i فإن قيمة الانحراف المعياري لا تتغير

5- إذا ضربنا كافة القيم بعدد ثابت وليكن C فإن التباين يساوي التباين للقيم الأصلية مضروبا في مربع المقدار الثابت والانحراف المعياري الجديد يساوي الانحراف المعياري للقيم الأصلية مضروباً في المقدار الثابت $S_Y = C S_X$

6- إذا قسمنا جميع القيم على مقدار ثابت C فإن الانحراف المعياري النتائج هو الانحراف المعياري للقيم الأصلية مقسوماً على المقدار الثابت C ويستعمل كل من التباين والانحراف المعياري في عدة مجالات أهمها مجال الاستثمار وهو قياس مخاطر الاستثمار في الأسهم من حيث عدم ثبات أرباحها فكلما زاد التباين والانحراف المعياري لأرباح سهم ما كلما زادت مخاطر الاستثمار فيه.

5. مقاييس التشتت النسبي:

1.5 معامل الاختلاف (Coefficient of Variation):

عزيزي الدارس، بمراجعة جميع مقاييس التشتت التي سبق لنا مناقشتها (المدى، الانحراف الربيعي، الانحراف المتوسط، التباين، الانحراف المعياري) نجد إنها مقاييس مطلقة محسوبة بدلالة وحدات القياس الأصلية المستقلة في قياس المتغيرات التي ندرسها ، سواء أكانت هذه الوحدات عبارة عن درجات أم سنوات أم غير ذلك . وإذا أردنا مقارنة درجة التشتت لمجموعتين أو صنفين متصلين بنفس المجموعة حال دون ذلك اختلاف وحدات القياس المستخدمة في الحالتين . فمثلا إذا أردنا مقارنة التشتت في أوزان مجموعة بالتشتت في أعمار نفس المجموعة أو مجموعة أخرى فنجد أن مقياس التشتت في المجموعة الأولى يكون بالكيلوجرام بينما يكون في المجموعة الثانية بالسنوات ولا يعقل أن تقارن بين كيلوجرام وسنوات، ونقول أن الكيلوجرام أكبر أو أقل من تشتت السنوات.ولإجراء هذه المقارنة لابد من التخلص من وحدات القياس، وذلك باستخدام مقياس نسبي يخلصنا من هذه الوحدات المختلفة ،ويمكن أن نجعل الانحراف المعياري مقياساً للتشتت بقسمته على الوسط الحسابي، فيكون خارج القسمة عبارة عن نسبة مجردة من التمييز يطلق عليها معامل الاختلاف أو معامل التغير ويحسب من العلاقة

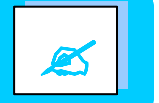
$$C.V = \frac{s}{\bar{x}}(100\%)$$

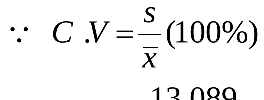
وهذا المقياس هو نسبة مئوية خالية من وحدات القياس، ولا يتأثر بكون القياسات كبيرة أو صغيرة مما يجعله صالحاً لمقارنة درجتي التجانس في عينتين من القياسات المستخدمة .

مثال(30.2):

مستخدماً بيانات المثال(27.2) أوجد معامل الاختلاف

الحل:

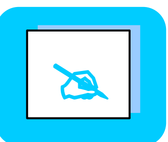




$$\because \quad C.V = \frac{s}{\bar{x}}(100\%)$$

$$\therefore C.V = \frac{13.089}{50} * (100) = 26.18\%$$

مثال(31.2):



بفرض أن الوسط الحسابي لأطوال مجموعة من الأشخاص 168سم وبانحراف معياري 9، في حين كان متوسط أوزانهم 76 كغم وبانحراف معياري 6 .

المطلوب: احسب معامل الاختلاف للطول والوزن، ثم حدد أيهما أكثر تشتتاً.

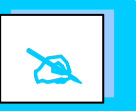
الحل

معامل الاختلاف للطول $C.V = \frac{9}{168} * (100) = 5.357\%$

معامل الاختلاف للوزن $C.V = \frac{6}{76} * (100) = 7.894\%$

أوضحت نتائج التحليل أن متغير الوزن هو أكثر تشتتاً من متغير الطول، أي أن الأشخاص أكثر تجانساً في أطوالهم عما هو في أوزانهم.

مثال(32.2):



إذا كان الوسط الحسابي لإنتاج العامل في المعمل A لإنتاج الأقمشة الجاهزة هو 1400 قطعة وبانحراف معياري 400، وكان الوسط الحسابي لإنتاج للعامل في المعملB لإنتاج الأقمشة الجاهزة هو 1300 قطعة وبانحراف معياري 300 خلال فترة شهر فحدد أي من المعملين أقل تشتت في إنتاج الأقمشة.

الحل: معامل الاختلاف لإنتاج العامل A =

$$C.V = \frac{S}{\bar{x}}(100\%) = \frac{400}{1400}(100\%) = 28.57\ \%$$

معامل الاختلاف لإنتاج العامل B =

$$C.V = \frac{S}{\bar{X}}(100\%) = \frac{300}{1300}(100\%) = 23.08\%$$

نلاحظ أن مقدار التشتت في المعمل B أقل مما هو عليه في المعمل A

2.5 القيم المعيارية:

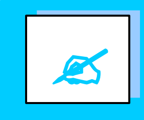
تعرف القيمة المعيارية بأنها عبارة عن خارج قسمة الفرق بين القيمة الفعلية والمتوسط الحسابي للبيانات على الانحراف المعياري وهي أحد مقاييس التشتت

النسبي ولتوضيح ذلك نفرض $x_1 , x_2 , ..., x_n$ هي مجموعة من القيم متوسطها $\bar{x}$ وانحرافها المعياري S ويرمز للقيمة المعيارية بالرمز Z وتحسب من العلاقة:

في حالة العينة
$$Z = \frac{X_i - \bar{X}}{S}$$

وتسمى المعادلة أعلاه تحويلة حيث يتم تحويل كل قيمة من قيم المتغير العشوائي X إلى قيمة معيارية Z وهي التي تستخدم كأساس لمقارنة القيم ببعضها على أساس الوحدات المعيارية.

مثال(33.2):



نفرض أن لدينا درجات طلاب في مادتي الفيزياء والكيمياء ونرغب في مقارنة درجتي أحد الطلبة في المادتين فإذا كانت درجته في الفيزياء 80 وفي الكيمياء 70 وإذا كان متوسط درجات الطلاب في الفيزياء 85 بانحراف معياري 5، وأن متوسط درجات الطلاب في الكيمياء هو 62 بانحراف معياري 6 فهل يعني ذلك أن تحصيله في الفيزياء أفضل من تحصيله في الكيمياء ؟

الحل :

نحسب الدرجة المعيارية للمادتين:

درجة الطالب المعيارية في الفيزياء Z_1

$$Z_1 = \frac{X_i - \bar{X}}{S} = \frac{80 - 85}{5} = -1$$

درجة الطالب المعيارية في الكيمياء Z_2

$$Z_2 = \frac{X_i - \bar{X}}{S} = \frac{70 - 62}{6} = 1.33$$

يلاحظ أن تحصيل الطالب في الكيمياء أفضل من تحصيله في الفيزياء على الرغم من أن درجته في الفيزياء أكبر من درجته في الكيمياء. لأن درجته في الكيمياء تبعد بمقدار 1.33 عن المتوسط.

تدريب (12)

لديك البيانات الآتية : 2، 3 ، 4 ، 6 ، 10.

أحسب كلاً من :

1) الانحراف المعياري لهذه البيانات .   2) معامل الاختلاف.

أسئلة التقويم الذاتي (4) :

1) إذا كان الانحراف المعياري لدرجات مجموعة من الطلبة هو 12 فما هو الانحراف المعياري بعد إضافة المدرس 6 درجات لكل طالب.

2) **أولاً**: أكمل الفراغات الآتية بما يناسبها :

1- يقصد بالتشتت من الناحية الإحصائية بأنه .....ويكون التشتت صفراً إذا .....

2- درجة التشتت إما أن تكون .....أو.....أو.....

3- من أهم استخدامات المدى استخدامه في خرائط مراقبة ....

4- التباين هو متوسط مربعات انحرافات القيم عن ....

ثانياً:

1-مجموعة من القيم كان تباينها يساوي 30 فكم يصبح التباين في الحالات التالية:

1- إذا ضربنا كل قيمة في 7 ؟

2- إذا قسمنا كل قيمة على 5 ؟

3-أضفنا 6 إلى كل قيمة في المجموعة ؟

2- من مجتمعين مختلفين أخذت عينتين فكانت النتائج كما يلي

العينة الأولى	العينة الثانية
$\sum_{i=1}^{40} Y_i = 390$	$\sum_{i-1}^{30} X_i = 270$
$\sum_{i=1}^{40} Y_i^2 = 3984$	$\sum_{1}^{30} X_i^2 = 2691$

أوجد :

1- الوسط الحسابي لكل عينة   2- التباين لكل عينة

3- الانحراف المعياري لكل عينة.

4- معامل الاختلاف لكل عينة. ثم حدد أيهما أكثر تشتت.

6 . مقاييس التماثل والالتواء:

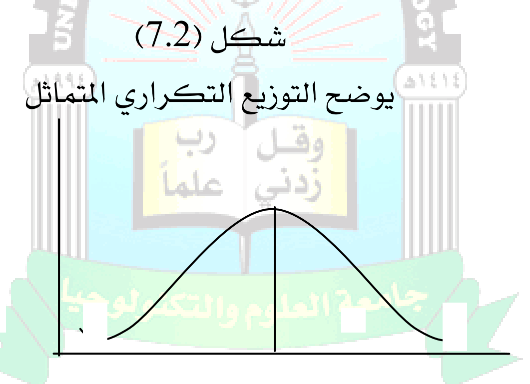
تذكر **عزيزي الدارس**، أن مفهوم التماثل والالتواء قد تم التعرض له في الوحدة الأولى، وكذلك في هذه الوحدة عندما ناقشنا سوياً العلاقة بين المتوسطات وذكرنا بأن التوزيع المتماثل تكون فيه المتوسطات الثلاثة متساوية. وفي حياتنا العملية نادرا ما نحصل على مثل هذا التماثل، إذ أن من الطبيعي أن يظهر الالتواء في التوزيع التكراري.

فيقال للتوزيع متماثل عندما يتطابق نصفيه عند محور عمودي شكل (7.2) وعندما لا يتطابق جانبي التوزيع يقال عنه توزيع ملتوي فعندما يكون الالتواء باتجاه اليمين يقال عنه توزيع موجب الالتواء شكل (8.2) حيث يصبح طرفه الأيمن أطول من طرفه الأيسر .

وعندما يكون الالتواء باتجاه اليسار فيدعى بالتوزيع السالب الالتواء ويصبح طرفة الأيسر أطول من طرفه الأيمن شكل(9.2)

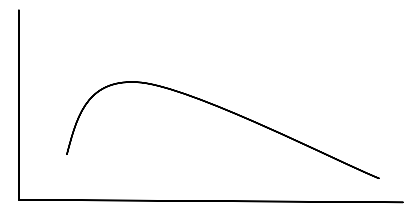
شكل (7.2)

يوضح التوزيع التكراري المتماثل



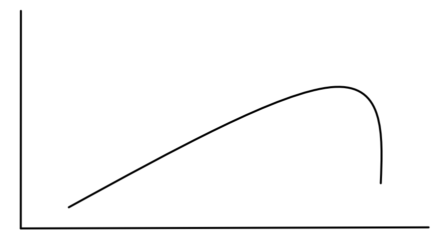
شكل (8.2)

يوضح التوزيع التكراري الموجب الالتواء



شكل (9.2)

يوضح التوزيع التكراري السالب الالتواء



وهناك أكثر من طريقة لحساب معامل الالتواء منها :

1.6 معامل (بيرسون) للالتواء

يعتبر **معامل بيرسون للالتواء** من أهم مقاييس الالتواء والتماثل ويرمز له بالرمز SK وصيغته هي:

$$Sk = \frac{\overline{X} - M_o}{S} \qquad (*)$$

حيث:

$\overline{X}$: الوسط الحسابي.

M_o : المنوال.

S : الانحراف المعياري.

ومن عيوبه شموله على المنوال الذي قد لا يكون موجوداً بين البيانات أو قد يكون للبيانات أكثر من منوال بالإضافة إلى صعوبة تحديده في حالة البيانات المبوبة، ولمعالجة هذا العيب نلاحظ أن **الوسيط في التوزيعات الملتوية يبعد حوالي الضعف عن المنوال** لذا يتم اشتقاق صيغة بديلة كما يلي:

$$M_o - M_d = 2(M_d - \bar{x})$$

$$M_o = 2M_d + M_d - 2\bar{x}$$

$$M_o = 3M_d - 2\bar{x}$$

نعوض في الصيغة (*) عن M_o

$$Sk = \frac{3(\bar{x} - M_d)}{S}$$

ويكون معامل (بيرسون) للالتواء موجب باتجاه اليمين إذا كان الوسط الحسابي اكبر من الوسيط والمنوال. ويكون سالباً باتجاه اليسار إذا كان $\bar{x}$ أقل من M_d ، M_o وعند تطابق المتوسطات الثلاثة فإن $Sk=0$.

وعموماً فإن قيمة Sk تقع بين $+3$, -3

مثال (34.2):



احسب معامل بيرسون لالتواء التوزيع إذا كان $\bar{X}=3.41$, $M_d=3.4$, $S=0.7$ ثم حدد درجة ونوع الالتواء

الحل:

$$\because\ Sk=\frac{3\,(\bar{X}-M_d)}{S}$$

$$\therefore\ Sk=\frac{3(3.41-3.4)}{0.7}=0.04$$

و من هذه النتيجة نستدل على أن درجة الالتواء في شكل التوزيع التكراري للبيانات بسيط ويكون موجباً في اتجاه اليمين.

تدريب (13)



إذا كانت الرواتب المحددة لخريجي الكليات عند بداية العمل في البنوك يبلغ متوسطها $\bar{X}$ وهو 2500 وبانحراف معياري s مقداره 700 فهل يعتبر الراتب 4000 راتباً اعتياديا يقع ضمن هذا المدى

تدريب (14)



علل : من عيوب معامل (بيرسون) للالتواء احتوائه على المنوال ؟ وكيف تعالج هذا العيب؟

7. مقاييس التفرطح أو التدبب Kurtosis or Peaked ness :

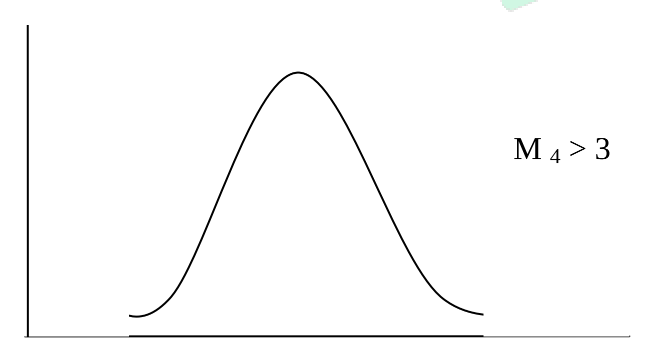
التفرطح هو مقياس يبين درجة علو قمة المنحنى مقارنة بالتوزيع الطبيعي، أي يقيس مدى تدبب قمة المنحنى التكراري وسوف ندرس فيما بعد التوزيع الاعتدالي ومن صفات هذا التوزيع أن تفرطحه يساوي 3 ويعتبر هذا التوزيع أساس المقارنة عند قياس التفرطح .وعليه فإن شكل التفرطح يأخذ عدة اشكال منها ما هو ذات أطراف واسعة نسبياً، وقمة ضيقة، ويطلق عليه عادة ذو القمة المرتفعة ويكون تفرطحه أكبر من 3 والذي يأخذ الشكل البياني رقم (10.2) وعندما يكون المنحنى معتدل التفرطح، يسمى بالتفرطح المعتدل ويكون مقياس تفرطحه يساوى 3 كما في الشكل رقم (11.2) أما عندما تكون قمة المنحنى مسطحاً فيدعى بالتوزيع، أو المنحنى المفرطح ويكون تفرطحه أقل من 3 كما هو الحال في الشكل البياني رقم (12.2) والذي عادة ما تكون أطرافه ضيقة.

وبحسب مقياس التفرطح من الصيغة $k_1 = \frac{m_4}{m_2^2}$

حيث m_4 هو العزم الرابع حول المتوسط ويحسب من الصيغة $m_4 = \frac{\sum (x_i - \bar{x})\, f_i}{n}$ و m_2^2 هو العزم الثاني حول المتوسط (التباين) أو هو القوة الرابعة للانحراف المعياري $M_d = 4.35$

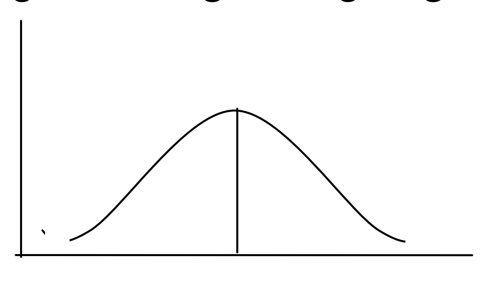
شكل (10.2):

يوضح التوزيع ذو التدبب المرتفع



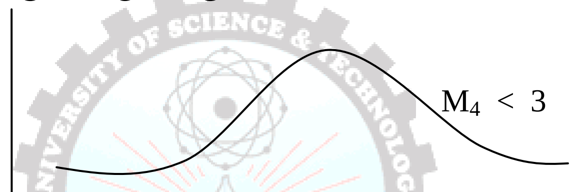
شكل (11.2)

يوضح التوزيع ذو التفرطح المعتدل (التوزيع الطبيعي)



شكل (12.2)

شكل يوضح التوزيع المفرطح



مثال (35.2):

احسب مقياس التفرطح في المثال (29.2).

الحل :

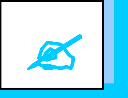
لحساب مقياس التفرطح من المثال السابق نلاحظ أن

$m_4 = 87.933 \quad , \quad m_2 = 6.0336$

$$\frac{m_4}{m_2^2} = \frac{87.933}{(6.0336)^2} = 2.415$$

وعليه يعتبر التوزيع مفرطحاً

تدريب (15)



عرف التفرطح ؟ ومتى يكون التوزيع معتدل؟

أسئلة التقويم الذاتي (5) :

1) أكمل الفراغات الآتية بما يناسبها :

1- يقال للتوزيع متماثل عندما يتطابق ....

2- يكون التوزيع سالب الالتواء عندما يكون الالتواء...........

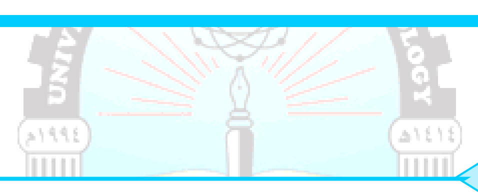
3- يكون معامل بيرسون للالتواء موجب باتجاه اليمين إذا كان .... أكبر من ....

2) ماذا نطلق على التوزيع التكراري الذي معامل التوائه يساوي صفراً.

3) احسب معامل بيرسون لالتواء التوزيع إذا كان، $\overline{X} = 4.3$ ، $M_d = 4.35$

$S=0.7$ ثم حدد درجة ونوع الالتواء .

4) ما نوع قيمة منحنى التوزيع التكراري إذا كان معامل التفرطح 2.415





نشاط

عزيزي الدارس : أقترح عليك أن تذهب إلى الجهاز المركزي للإحصاء وتطلب بيانات عن السكان في الجمهورية اليمنية ومن خلالها تقوم بحساب المتوسط الحسابي لعدد أفراد الأسرة ، وكذلك تحدد نسبة الإناث والذكور إلى عدد السكان في اليمن .

8. الخلاصة:

تناولت هذه الوحدة قدراً كبيراً من الإحصاء الوصفي حيث اشتملت الوحدة على عدد من المقاييس الإحصائية ، وسألخص لك أهم الموضوعات التي وردت في مادة الوحدة، وأرجو أن تجدها مفيدة لك:

1- مقاييس النزعة المركزية :

- **النزعة المركزية**: هي ميل معظم المفردات للتجمع حول قيمة واحدة تسمى القيمة المتوسطة

- **الوسط الحسابي**: هو مجموع قيم البيانات مقسوما على عددها

- **الوسيط**: هو تلك القيمة التي تتوسط مجموعة القيم بحيث يسبقها نصف القيم ويتلوها النصف الآخر

- **المنوال**: القيمة الأكثر شيوعاً أو تكراراً في البيانات.

- **الفئة المنوالية**: هي الفئة التي يقابلها أكبر تكرار.

- من أهم خواص الوسط الحسابي تأثره بالقيم المتطرفة، وأن مجموع انحرافات القيم عن وسطها الحسابي يساوي صفراً. بينما يتأثر الوسيط بالقيم الوسطي، ويتأثر المنوال بالتكرارات

وفي حالة التوزيع المتماثل تكون المتوسطات الثلاثة متساوية(الوسط الحسابي= الوسيط = المنوال)

- **العشيريات**: هي التي تقسم البيانات إلى عشرة أقسام متساوية بعد ترتيبها تصاعدياً أو تنازلياً

- **الرُّبيع** : هو الذي يقسم البيانات إلى أربعة أقسام متساوية بعد ترتيبها تصاعدياً أو تنازلياً

- **المئين** : يعني تقسيم البيانات إلى مائة قسم متساوٍ بعد ترتيبها تصاعدياً أو تنازلياً

2- مقاييس التشتت

- **التشتت**: هو مدى التقارب أو التباعد بين المفردات حيث تكون درجة التشتت صفر(إذا تساوت جميع القيم)أو ضعيفة(قريبة من الصفر) أو عالية، ومقدار التشتت موجب إذ لا معنى للتشتت السالب.

- **المدى**: هو الفرق بين أكبر وأصغر قيمة في البيانات، ويستخدم في خرائط مراقبة جودة الإنتاج

- **التباين** : هو متوسط مجموع مربعات انحرافات القيم عن الوسط الحسابي، إلا أنه يعتمد على قيم غير أصلية وهي مربعات انحرافات القيم عن الوسط الحسابي.

- **الانحراف المعياري** : هو الجذر التربيعي للتباين وهو أكثر مقاييس التشتت استخداماً

3- مقاييس التشتت النسبي والتي تستخدم لأغراض المقارنة بين مجموعتين أو أكثر ومن هذه المقاييس:

- **معامل الاختلاف**: الذي يساوي الانحراف المعياري مقسوما على الوسط الحسابي مضروباً في مائة وهو نسبة مئوية خالية من وحدات القياس.

- **القيم المعيارية** وهي: عبارة عن الفرق بين القيمة الفعلية ووسطها الحسابي مقسوما على الانحراف المعياري للبيانات.

كما تناولت الوحدة التماثل والالتواء حيث يكون المنحنى متماثل إذا كان طرفه الأيمن يساوي طرفه الأيسر، أما إذا كان طرفه الأيمن أطول من طرفه الأيسر فيقال أن المنحنى ملتوي باتجاه اليمين موجب الالتواء ، وإذا كان طرفه الأيسر أطول فيقال أن المنحنى ملتوي باتجاه اليسار سالب الالتواء.

9. لمحة مسبقة عن الوحدة الثالثة :

بعد دراستك للوحدة الثانية (المقاييس الإحصائية) أصبحت قادرا على عرض وتلخيص البيانات على شكل توزيع تكراري ، و تحديد فيما إذا كان التوزيع ملتوي أم لا ، وتحدد طبيعة التوزيع بطرق مختلفة من خلال حساب بعض المقاييس الإحصائية.

وسنوسع نطاق الدراسة في الوحدة الثالثة من خلال دراستنا للعلاقات بين متغيرين وليس متغير واحد ، لأن دراسة متغير لوحدة لا يكفي لعرض الظواهر المختلفة وإنما علاقة المتغير بمتغير آخر وهو ما يعرف بالارتباط وسنوضح كيفية حساب معامل الارتباط الخطي البسيط ومعامل ارتباط الرتب ، وستزداد معرفتك من خلال تناولنا لموضوع الانحدار الخطي البسيط والذي تستخدم نماذجه للتنبؤ المستقبلي للكثير من الظواهر، وسنتعرض بالدراسة لمعامل الاقتران الخطي. ونختم هذه الوحدة بدراسة السلاسل الزمنية بهدف الكشف عن التغيرات التي تطرأ على الظواهر تحت الدراسة خلال فترة زمنية محددة وتفيد في التنبؤ المستقبلي.

جامعة العلوم والتكنولوجيا



10. إجابات التدريبات:

تدريب (1) :

1- مقاييس النزعة المركزية هي :

- الوسط الحسابي، الوسيط، المنوال، الوسط التوافقي، الوسط الهندسي ،

2- يعرف الوسط الحسابي بأنه مجموع قيم المجموعة مقسوماً على عددها.

ويعرف بأنه ذلك المقياس الإحصائي الوصفي الذي إذا حسبنا انحرافات مفردات المجموعة عنه لكان مجموع هذه الانحرافات يساوي الصفر.

تدريب (2):

يحسب الوسط الحسابي من الصيغة

$$\overline{X} = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$\overline{X} = \frac{25+9+23+22+14+10+12+18+20}{9}$$

$$\overline{X} = \frac{153}{7} = 17$$

تدريب (3) :

بعض خواص الوسط الحسابي:

1. مجموع انحرافات القيم عن وسطها الحسابي يساوي صفر.
2. تتأثر قيمة الوسط الحسابي بالقيم المتطرفة (القيم الصغيرة جداً أو الكبيرة جداً)
3. الوسط الحسابي يأخذ جميع القيم بالاعتبار..
4. الوسط الحسابي محدد جبرياً بدقة ويتمتع بخواص جبرية لا يتمتع بها غيره من المقاييس.
5. يعتبر الوسط الحسابي غير ممكن الحساب في حال البيانات النوعية.
6. لا يمكن حساب الوسط الحسابي من الجداول التكرارية المفتوحة
7. لا يمكن إيجاد الوسط الحسابي بالطريقة البيانية إلا في حالة التوزيع
8. التكراري المتماثل.

تدريب (4):

نرتب البيانات تصاعدياً

11، 14، 22، 25، 28، 35

وبما أن عدد المفردات زوجياً وعددها يساوي 6

نوجد رتبة الوسيط الأول من الصيغة $\frac{n}{2}=\frac{6}{2}=3$

فتكون قيمة الوسيط الأول هي القيمة ذات الترتيب الثالث وهي 22

نحدد رتبة الوسيط الثاني من الصيغة $\frac{n}{2}+1=\frac{6}{2}+1=4$

فتكون قيمة الوسيط الثاني هي القيمة ذات الترتيب الرابع وهي تساوي 25

ونحصل على الوسيط من الصيغة $\frac{22+25}{2}=23.5$

تدريب (5) :

1- أهم خواص الوسيط :

1) يقع الوسيط في أي توزيع تكراري غير متماثل بين الوسط الحسابي والمنوال

2) يتأثر الوسيط بالقيم الوسطى ولا يتأثر بالقيم المتطرفة.

3) يمكن حسابه إذا كانت إحدى الفئات مفتوحة من أسفل أو من أعلى، كما يمكن حسابه إذا كانت الفئات غير متساوية الطول.

4) لا يتمتع بأي خواص جبرية بحيث يمكن الاستفادة منه في حسابات لاحقة.

2-يمكن تحديد موقع الوسيط من خلال قسمة مجموع التكرارات على أثنين من الصيغة $\frac{\sum f_i}{2}$

تدريب (6) :

1- المنوال: هو القيمة الأكثر شيوعاً أو تكراراً في البيانات

2- أهم **خواص المنوال:**

1) لا يتأثر المنوال بالقيم المتطرفة في التوزيع التكراري بل يتأثر بالتكرارات

2) يمكن حسابه من الجداول التكرارية المفتوحة

3) يعتمد في حسابه على جزء من البيانات

4) قد يوجد أكثر من منوال للبيانات وفي هذه الحالة يقال أن البيانات غير متجانسة

5) يستخدم في حالة البيانات النوعية

3- نلاحظ أن العدد 105 تكرر ثلاث مرات وهو أكثر الأعداد تكراراً وعليه فإن المنوال يساوي 105

تدريب(7) :

بما أن المنوال هو: هو القيمة الأكثر شيوعاً أو تكراراً في البيانات

فإننا نلاحظ أن القيمة 13.5 تكررت ثلاث مرات

فيكون المنوال هو القيمة 13.5

تدريب(8):

نعلم أنه في حالة التوزيعات الشبه ملتوية أن:

الوسط الحسابي – المنوال = 3 (الوسط الحسابي – الوسيط)

أو باستخدام الرموز

$$\bar{X} - M_o = 3(\bar{x} - M_d)$$

$$\bar{X} - 156.647 = 3(\bar{x} - 155.536)$$

$$\bar{X} - 156.647 = 3\bar{x} - 466.608$$

$$\bar{X} - 3\bar{x} = 156.647 - 466.608$$

$$-2\bar{x} = -309.961$$

$$\bar{X} = \frac{-309.961}{-2} = 154.981$$

تدريب(9):

يتم تحديد موقع الربيع أو العشير أو المئين في حالة البيانات غير المبوبة من الصيغ الآتية:

موقع الربيع $\frac{(i)(n+1)}{4} \quad , \ i=1,2,3$

موقع العشير $\frac{(i)(n+1)}{10} \quad , \ i=1,2,3,.....,9$

موقع المئين $\frac{(i)(n+1)}{100} \quad , \; i=1,2,3,.......,99$

أما في حالة التوزيعات التكرارية فإننا سوف نحدد موقع كل مقياس بنفس الطريقة للبيانات غير المبوبة إلا أننا سوف نستخدم $\sum f_i$ بدلاً من n وهي لا تعتبر من مقاييس النزعة المركزية

تدريب(10):

المدى = أكبر قيمة - أصغر قيمة

المدى = 25 – 10 = 15

تدريب(11):

1- من الصعب التعامل مع الانحراف المتوسط رياضياً لأنه يهمل الإشارة الجبرية بدون مبرر علمي يعتمد عليه.

2-

$$\bar{X} = \frac{\sum X_i}{n} = \frac{175}{5} = 35$$

$$M.D = \frac{\sum |X_i - \bar{X}|}{n} = \frac{40}{5} = 8$$

X_i	$X_i - \bar{X}$	$\lvert X_i - \bar{X} \rvert$
35	0	0
20	-15	15
30	-5	5
40	5	5
50	15	15
$\sum$	0	40

تدريب(12):

نكون الجدول التالي:

i	X_i	$X_i-\bar{X}$	x^2	$(X_i-\bar{X})^2$
1	2	-3	4	9
2	3	-2	9	4
3	4	-1	16	1
4	6	1	36	1
5	10	5	100	25
$\sum$	25	--	165	40

$$\bar{X}=\frac{\sum_{i=1}^{n}X_i}{n}=\frac{2+3+4+6+10}{5}=\frac{25}{5}=5$$

$$S=\sqrt{\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{n}}$$

$$=\sqrt{\frac{40}{5}}=\sqrt{8}=2.83$$

أو باستخدام العلاقة الآتية

$$S=\sqrt{\frac{\sum_{i=1}^{n}X_i^2}{n}-\left(\frac{\sum_{i=1}^{n}X_i}{n}\right)^2}=\sqrt{33-25}=2.83$$

$$C.V=\frac{S}{\bar{X}}\ X/\omega=\frac{2.83}{5}\ X100=56..5$$

تدريب(13):

نحسب القيمة المعيارية

$$Z=\frac{x-\mu}{\sigma}$$

$$=\frac{4000-2500}{700}=2.143$$

وهذا يعني أن الراتب السنوي الذي قيمته 4000 ريال يزيد بمقدار 2.143 عن الوسط الحسابي وهو بذلك راتباً اعتيادياً يقع ضمن المدى المطلوب.

تدريب(14):

من عيوب معامل (بيرسون) للالتواء شموله على المنوال للأسباب:

1- قد لا يوجد للبيانات منوال

2- قد يوجد للبيانات أكثر من منوال

3- صعوبة تحديده في حالة البيانات المبوبة.

ولمعالجة هذا العيب نلاحظ أن الوسيط في التوزيعات الملتوية يبعد حوالي الضعف عن المنوال لذا يتم اشتقاق صيغة بديلة

$$M_o - M_d = 2\left(M_d - \overline{X}\right)$$

$$M_o = 2M_d + M_d - 2\overline{X}$$

$$M_o = 3M_d - 2\overline{X}$$

نعوض عن M_o في الصيغة

$$SK = \frac{\overline{X} - M_o}{S}$$

نجد أن

$$SK = \frac{\overline{X} - 3M_d + 2\overline{X}}{S}$$

$$\therefore SK = \frac{3\left(\overline{X} - M_d\right)}{S}$$

تدريب(15):

يقصد بالتفرطح هو درجة تدبب منحنى التوزيع، أي يقيس مدى تدبب قمة المنحنى التكراري، ويكون التوزيع معتدل عندما يكون تفرطحه يساوي 3.

11. قائمة المصطلحات

مقاييس النزعة المركزية (Measures Central Tendency)

- **الوسط الحسابي (The Arithmetic Mean)**: هو مجموع القيم مقسوماً على عددها

- **الوسيط (The Median)**: وهو القيمة التي تتوسط مجموعة القيم بحيث يسبقها نصف عدد القيم ويتلوها النصف الآخر . بعد ترتيبها ترتيباً تنازلياً أو تصاعدياً.

- **المنوال (The Mode)** : **وهو القيمة الأكثر تكراراً**

- **مقاييس التشتت (Measures of Variation)**: المدى، المدى الربيعي، والانحراف المتوسط، والتباين والانحراف المعياري وتسمى مقاييس التشتت المطلقة.

التشتت : هو مدى تقارب أو تباعد قيم مفردات رقمية لظاهرة ما حول أي قيمة متوسطة لتلك البيانات (المفردات).وهذه القيمة قد تكون الوسط الحسابي أو الوسيط أو المنوال.

- **المدى (The Range)**: وهو الفرق بين أكبر وأصغر قيمة في البيانات

التباين والانحراف المعياري (Variance and Standard Deviation)

- **التباين (The variance)** : متوسط مجموع مربعات انحرافات القيم عن وسطها الحسابي ويرمز له بالرمز S^2 في حالة حسابه للعينات، أما في حالة حسابه للمجتمع الإحصائي فنرمز له بالرمز σ^2 .

- **الانحراف المعياري (Standard Deviation)**: هو أحد مقاييس التشتت الهامة وأكثرها شيوعاً أو استخداماً في علم الإحصاء ويرمز له بالرمز S إذا كانت البيانات تمثل عينة من مجتمع و σ إذا كانت البيانات تمثل المجتمع الإحصائي، و يعرف بأنه الجذر التربيعي للتباين

- **معامل الاختلاف (Coefficient of Variation)** وهو أحد مقاييس التشتت النسبي ويستخدم للمقارنة بين مجموعتين من البيانات تختلفان في وحدات القياس.

12. المراجع العربية والأجنبية

1.12 المراجع العربية

1- أبو صالح، محمد صبحي وعوض، عدنان محمد.(1995): مقدمة في الإحصاء . الطبعة الثانية، مركز الكتاب الأردني.

2- البشير ،زين العابدين عبد الرحيم وعودة، أحمد عودة عبد المجيد.(1997): الاستدلال الإحصائي. منشورات جامعة الملك سعود .

3- البلداوي، عبد الحميد .(1997) : الإحصاء للعلوم الإدارية والتطبيقية . الطبعة العربية الأولى، دار الشروق، عمان : الأردن.

4- توفيق، عبد الجبار. (1983): التحليل الإحصائي في البحوث التربوية والنفسية والاجتماعية الطرق اللامعلمية . الطبعة الأولى، مؤسسة الكويت : الكويت .

5-الجهاز المركزي للإحصاء.(2002):كتاب الإحصاء السنوي 2001م صنعاء: اليمن.

6-حبيب، مجدي عبد الكريم .(2001): الإحصاء اللابارامتري الحديث في العلوم السلوكية. الطبعة الأولى. مكتبة النهضة المصرية، القاهرة : مصر

7- رمضان، زياد .(1997) : مبادئ الإحصاء الوصفي والتطبيقي والحيوي.الطبعة الرابعة، دار وائل للنشر، عمان : الأردن.

8- العواد، منذر حسين .(2003): مبادئ الإحصاء . الطبعة الأولى، مركز الأمين للنشر، صنعاء : اليمن.

9- فراج، عبد الحميد .(1975): الأسس الإحصائية للدراسات السكانية . دار النهضة العربية، القاهرة : مصر.

10- كنجو، أنيس إسماعيل .(2000) : الإحصاء والاحتمال . الطبعة الأولى، مكتبة العكيبان، الرياض : السعودية .

11- مصطفى، مدني دسوقي.(1968): مبادئ في نظرية الاحتمالات والإحصاء الرياضي وتطبيقاتها في الاستنتاج الإحصائي . دار النهضة العربية، القاهرة : مصر .

12- المنصوب، محمد عبد الكريم .(1998): مفاهيم أساسية في الإحصاء . الطبعة الأولى، منشورات دار الخبرة، صنعاء: اليمن.

2.12 المراجع الأجنبية

1- Anderson , D . B., Sweeney , D . J. and Williams , T . A , (1991) Introduction to Statistics : Concepts and Application , 2nd end , West Publishing Company .

2- Cramer, H.(1961): Mathematical Methods of Statistics. Princeton: Princeton University press.

3- Dixon & Massey , Introduction to Statistical Analysis.

4- Draper N . and Smith H ., Applied Regression Analysis ,John Wiley and sons Inc ., London 1990

5 -Hayslett , H . T. (J r) , Statistics Made Simple , Made simple Books , Doubleday and Co ,. Inc .Garden City , New York , 1968 .

6- Kendall, M. and Stuart , A.(1977): Advanced Theory of statistics, vol 1, 4th ed . London : Charles Griffin & Company.

7- Sendecor G . and Cochran G ., Statistical Methods , 7th Ed ., The Iowa State Universtiy Press , U . S . A . 1980.

جامعة العلوم والتكنولوجيا
وقل رب زدني علماً

13. التعيينات

السؤال الأول:

أختر الإجابة الصحيحة من بين الإجابات الواردة بوضع دائرة حول رقم الإجابة الصحيحة:

1- من خصائص الوسط الحسابي أنه:

1- يتأثر بالقيم الوسطى في المجموعة

2- يتأثر بالقيم المتطرفة في المجموعة

3- مقياس غير محدد بدقة في الجداول التكرارية

2- لمجموعة من القيم وسطها الحسابي 21 وانحرافها المعياري 3 فإذا ضربنا كل قيم المجموعة في 3 فإن الوسط الحسابي يساوي:

1- 32   2- 31.5   3- 63

3- الانحراف المعياري في 2 يساوي:

أ- 9   ب- 4.5   جـ- 3

مستعيناً ببيانات الجدول التكراري التالي أجب عن الفقرات التالية للجدول:

الفئات	f_i	x_i	$f_i\ x_i$	X^2_i	f_ix^2	تكرار متجمع صاعد
10-14	2	12	24	144	288	2
15-19	6	17	102	189	1734	8
20-24	8	22	176	484	3872	16
25-29	5	27	135	729	3645	21
30-34	5	32	160	1024	5120	26
35-39	4	37	148	1369	5476	30
مجموع	30		745		20135	

4-الوسط الحسابي يساوي:

1- 24.11　　2- 24.83　　3- 24.38

5– المنوال يساوي:

1- 23　　2- 20　　3- 22

6- الانحراف المعياري يساوي:

1- 7.39　　2- 7.93　　3- 8.39

7- التباين يساوي:

1- 54.638　　2- 53.638　　3- 54.10

8- الربيع الثالث يساوي:

1- 32.31　　2- 32　　3- 31.5

9- للبيانات التالية 5,4,3,8,7 فإن الوسط الحسابي هو:

1- -5.4　　2- 5.4　　3- 5.2

10- من عيوب معامل (بيرسون) احتوائه على المنوال وذلك بسبب :

1- قد لا يوجد منوال للبيانات　　2-قد يوجد للبيانات أكثر من منوال

3- صعوبة استخراجه من الجداول التكرارية

4- كل ما ذكر

11- يعرف الانحراف المتوسط بأنه:

1-معدل الانحرافات المطلقة لكافة المعطيات x_i عن قيمة المتوسط الحسابي

2-القيمة التي تتحرف بها نقاط الربيع الأول والثالث عن الوسيط

3-متوسط مربعات انحرافات القيم عن وسطها الحسابي

12- إذا كان المتوسط الحسابي = 12.5 والوسيط =10 والانحراف المعياري =0.9

فإن معامل (بيرسون) للالتواء هو:

1- 2.778　　　2- 2.078　　　3- 3.278

السؤال الثاني:

ضع علامة (√) أمام الجمل الصحيحة وعلامة (X) أمام الجمل الخاطئه فيما يأتي:

1( ) يمكن حساب الوسط الحسابي للجداول المفتوحة

2( ) يتأثر المنوال بالقيم المتطرفة

3( ) في حالة التوزيع السالب الالتواء يكون المنوال أصغر من كل من الوسط والوسيط

4( ) نقطة الضعف في التباين هي اعتماده على قيم غير أصلية

5( ) للبيانات التي تحتوي على قيم متطرفة يفضل استخدام نصف المدى الربيعي

6( ) يمكن إيجاد الوسط الحسابي بالطرق البيانية إذا كان المنحنى غير متماثل

7( ) إذا وجد أكثر من منوال فإنه يقال أن البيانات غير متجانسة

8( ) الربيع الأدنى يدعى بالمئين الخامس والعشرون

9( ) نقطة الضعف في الانحراف المعياري هي اعتماده على مربعات الانحرافات في قياس درجات التشتت

10( ) من خصائص الانحراف المعياري عدم تأثره في حالة ضرب قيم المجموعة بمقدار ثابت

السؤال الثالث:

أي مقاييس النزعة المركزية أفضل ؟ ولماذا ؟

UNIVERSITY OF SCIENCE & TECHNOLOGY
١٤١٤هـ ١٩٩٤م
وقل رب زدني علما
جامعة العلوم والتكنولوجيا

3 الوحدة الثالثة

الإرتباط والإنحدار والسلاسل الزمنية


UNIVERSITY OF SCIENCE & TECHNOLOGY
١٩٩٤م
١٤١٤هـ
وقل رب
زدني علماً

محتويات الوحدة



1- المقدمــة:

1.1 التمهيـد:

عزيزي الدارس، مرحبا بك إلى دراسة هذه الوحدة التي تنقسم إلى ثلاثة أقسام رئيسة هي الارتباط والانحدار والسلاسل الزمنية. وقد سبق أن تعرفنا في الوحدتين السابقتين على بعض أساليب التحليل الإحصائي والتي يمكن استخدامها لتحليل متغير واحد مثل أطوال مجموعة من الأشخاص أو أوزانهم أو الدخل الشهري وقيمة استهلاك الكهرباء أو الماء . فرأينا كيف يمكن أن نجمع البيانات ونلخصها ونحسب لها مقاييس النزعة المركزية، والتشتت، وكيف نحدد طبيعة التوزيع.

وفي هذه الوحدة سوف نتعرف على أساليب تحليل جديدة نستطيع باستخدامها تحليل متغير من خلال علاقته بمتغير آخر أو بعدة متغيرات أخرى والأمثلة على هذه الظواهر أو المتغيرات التي يوجد بينها علاقة كثيرة مثل: دخل الأسرة وإنفاقها، الغلة الزراعية وكمية السماد المستخدم. ونستطيع من خلال دراسة علاقة متغير ما بمتغير أو متغيرات أخرى أن نحدد وجود العلاقة بين المتغيرات وهذا الأسلوب يدعى تحليل الارتباط

وتستطيع **عزيزي الدارس** أيضاً، معرفة تأثير أحد المتغيرين على الآخر أو تأثير عدة متغيرات على متغير واحد من خلال دراسة الانحدار، ويتم تحليل الارتباط والانحدار من خلال إجراء عدد من الحسابات يسبقها التأكد من وجود علاقة بين المتغيرات.

لاحظنا في الأقسام السابقة أن العلاقات التي تمّت دراستها بين المجموعات كانت دون اعتبار لعامل الزمن، فكيف يمكن للمؤسسات المختلفة أن تتخذ القرارات المتعلقة بأوجه نشاطاتها الإدارية والاقتصادية والتخطيط لها بدون عامل الزمن. إذا كان التنبؤ يعتمد على ما حدث في الماضي ويحدث في الحاضر فإن التخطيط واتخاذ القرارات يعتمد أساساً على التنبؤ المستقبلي بما يحتمل وقوعه من أحداث ؛ فإن الزمن يعد من أهم العوامل التي لا يمن تجاهلها عند اتخاذ القرار في كل مجالات الحياة، وإذا كان هدف دراسة الانحدار هو التنبؤ المستقبلي فإن السلاسل الزمنية تعد من أهم أساليب الاستدلال حول المستقبل بناء على أحداث الماضي والحاضر والتي توضح التغير الذي يحدث في قيم متغير ما كدالة في الزمن سواء كان هذا التغير منتظما أو غير منتظم

2. 1 -أهداف الوحدة:

بعد فراغك **عزيزي الدارس** من دراسة هذه الوحدة يجب أن تكون قادرا على أن:

1-توضح مفهوم الارتباط .

2-تستخدم شكل الانتشار كوسيلة تمهيدية لمعرفة العلاقة بين متغيرين كميين من حيث كونها طر دية أو عكسية.

3- تحسب معامل الارتباط الخطي البسيط بين متغيرين وتفسّر نتيجته .

4- تحسب معامل ارتباط الرتب بين متغيرين وصفيين .

5- تفرق بين معامل ارتباط (بيرسون) ومعامل ارتباط (سبيرمان).

6- تقارن بين مقاييس الارتباط المختلفة، و متى يستخدم كل مقياس

7- تبين مفهوم الانحدار .

8- توضح المقصود بطريقة المربعات الصغرى .

9- تحسب معاملات خط انحدار متغير تابع على متغير مستقل.

10- تقدر قيمة المتغير التابع عند معرفة قيمة المتغير المستقل.

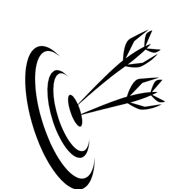
11- تحسب معامل التوافق وتفسر نتيجته.

12- تُعرّف السلسلة الزمنية.

13- توضح الهدف من الدراسة الإحصائية للسلاسل الزمنية.

14- تذكر العناصر الأساسية للسلسلة الزمنية.

15- تُبين كيفية استبعاد اثر التغيرات الموسمية.



3.1 أقسام الوحدة

قسمت هذه الوحدة إلى ثلاثة أقسام رئيسة:

القسم الأول: منها يتناول ظاهرة الارتباط من خلال دراسة العلاقة بين متغيرين كميين للمفردة الإحصائية الواحدة وكيفية توضيح العلاقة بينهما. من خلال شكل الانتشار كوسيلة مبدئية لتحديد طبيعة العلاقة بين هذين المتغيرين لتحديد إذا كانت هذه العلاقة طردية أو عكسية، خطية أم غير خطية وهذا ما يحقق

الهدفين الأول والثاني، أما الهدفين الثالث والرابع فيتحققان من خلال تعريف معامل الارتباط وكيفية حسابه، وتفسير دلالة قيمة معامل الارتباط. ويتحقق الهدفين الخامس والسادس من خلال حساب قيمة معامل الارتباط بين متغيرين كميين (بيرسون) أو أحدهما على الأقل وصفي من خلال رتبتهما وليس قراءاتهما.

أما في القسم الثاني: فسوف تتعرف عزيزي الدارس، على مفهوم الانحدار وكيفية بناء نموذج الانحدار الخطي باستخدام طريقة المربعات الصغرى، وتقدير معالم نموذج الانحدار (ميل الانحدار والمقدار الثابت) بهدف التنبؤ المستقبلي وهذا يحقق الأهداف السابع والثامن والتاسع، وفي ضوء نموذج الانحدار (خط الانحدار) تستطيع تقدير قيمة المتغير التابع بمعلومية قيمة المتغير المستقل وهذا يحقق الهدف العاشر .أما الهدف الحادي عشر فيرتبط بمعرفتك بطريقة حساب معامل التوافق والتعليق على نتيجته.

ويرتبط القسم الثالث: من هذه الوحدة بالسلاسل الزمنية والذي تتعرف فيه على مفهوم السلسلة الزمنية والهدف الإحصائي من دراستها وهذا يحقق الهدفين الثاني عشر والثالث عشر. أما الهدف الرابع عشر فيرتبط بدراسة عناصر السلسلة الزمنية وأهم ما يميز كل عنصر من هذه العناصر وأما الهدف الخامس عشر والأخير فيتحقق من خلال دراسة كيفية تعيين الاتجاه العام للسلسلة الزمنية واستبعاد اثر(الاتجاه العام، والتغيرات الموسمية).

4. 1 - القراءات المساعدة:

عزيزي الدارس، لكي ترسخ في ذهنك المفاهيم التي سترد في هذه الوحدة، حاول الانتفاع ما أمكن بالقراءات الآتية نظراً لاتصالها المباشر بموضوع هذه الوحدة. ولا شك أن انتفاعك بها سيعمق فهمك واستيعابك للموضوع ويوسع مداركك وآفاقك.



1- الهيتي، صلاح الدين حسين.(2004) : الأساليب الإحصائية في العلوم الإدارية .الطبعة الأولى، دار وائل للطباعة والنشر . عمان: الأردن.

2- أبو صالح، محمد صبحي وعوض، عدنان محمد.(1995): مقدمة في الإحصاء. الطبعة الثانية، مركز الكتاب لأردني.

3- البلداوي، عبد الحميد .(1997): الإحصاء للعلوم الإدارية والتطبيقية . الطبعة العربية الأولى، دار الشروق، عمان: الأردن.

5. 1 - الوسائط التعليمية المساندة:

قبل أن تبدأ بدراسة هذه الوحدة تأكد أن حقيبتك التعليمية تحتوي على الوسائط الآتية:



1. شرائح للقوانين الخاصة بالموضوعات التي تناولتها الوحدة.
2. أوراق رسم بياني لرسم شكل الانتشار.
3. الـ CD. R الخاص بالمادة.
4. جهاز حاسوب.

6. 1 - ما تحتاج إليه لدراسة الوحدة:

قبل أن تبدأ دراسة هذه الوحدة تأكد بأنك قد هيأت المكان الهادئ ، والمريح للدراسة، ولديك آلة حاسبة ، وقلم ، ومسطرة وبعض الأوراق للكتابة.

وفي أثناء دراستك للوحدة حاول الإجابة على أسئلة التقويم الذاتي ، لأنها تساعدك في مراجعة وتلخيص الوحدة، ولا تغفل التدريبات التي يجب عليك القيام بها فهي سترسخ الأفكار المعروضه في الوحدة وتمنحك الفرصة لاختبار تعلمك ومدى استيعابك وفهمك لمحتوى هذه الوحدة. ولا تتردد في الاتصال بمرشدك الأكاديمي كلما اقتضت الحاجة إلى ذلك فهو سيكون سعيداً بذلك وسيرحب بك كثيراً.



2. الارتباط (Correlation):

من خلال دراستنا في الوحدتين السابقتين تعرفنا على بعض الأساليب الإحصائية والتي تستخدم لتحليل متغير واحد مثل أعمار مجموعة من الأشخاص أو أوزانهم أو أطوالهم. وفي هذا الفصل سوف ندرس أساليب تحليل إحصائية جديدة تستخدم لتحليل متغير ما من خلال علاقته بمتغير آخر أو بعدة متغيرات مثل دراسة العلاقة بين الدخل والإنفاق للأسرة والعلاقة بين الدخل والحالة التعليمية للشخص...الخ. ومن خلال دراسة العلاقة بين متغير وآخر نتمكن من تحديد وجود العلاقة بين المتغيرات ونوعها ومتانتها وهذا الأسلوب يدعى تحليل الارتباط.

ويمكن تقسيم العلاقات بين الظواهر إلى قسمين هما:

أ- **علاقات تابعية**: وذلك عندما يكون مقابل كل قيمة للمتغير المستقل قيمة للمتغير التابع مثل العلاقة بين مساحة المربع وطول ضلعه وعلاقة مساحة الدائرة بنصف القطر.

ب- **علاقات ارتباطيه**: وذلك عندما يكون مقابل كل قيمة للمتغير المستقل قيمة تقريبية أو احتمالية للمتغير التابع ومن أمثلة العلاقات الارتباطية : مبيعات أفران الغاز، واسطوانات الغاز حجم المنشأة وإنتاجها، مؤهل العامل وأجره وعدد الوحدات المنتجة ونصيب الوحدة الواحدة من التكاليف الثابتة. وسوف نركز على النوع الثاني لأن النوع الأول سبق للطالب دراستها في الثانوية. مع ملاحظة أن العلاقة بين المتغيرات أو الظواهر قد تكون إما خطية وتمثل بخط مستقيم أو منحنية (غير خطية) وتمثل بمنحنى غير خطي وسنكرس دراستنا لدراسة الحالة التي تكون فيها علاقة الارتباط خطية بين الظاهرتين، حتى نتمكن فيما بعد أن نحدد قوة العلاقة واتجاهها.

والارتباط هو: ظاهرة إحصائية تطلق على العلاقة بين متغيرين، وهناك الكثير من المسائل العملية التي تكون فيها معرفة العلاقة بين متغيرين أو أكثر مهمة، ومن الأمثلة على ذلك:

1- دراسة العلاقة بين دخل الفرد وعدد ساعات العمل

2- دراسة العلاقة بين أوزان وأطوال مجموعة من الأطفال في عمر معين.

3- معرفة العلاقة بين عدد الساعات التي يدرسها الطالب ومعدله التراكمي في الجامعة.

1.2 معامل الارتباط (Coefficient of Correlation):

يعرف معامل الارتباط : بأنه مقياس لدرجة قوة العلاقة الخطية واتجاهها بين متغيرين.

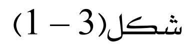
كما يقصد بمعامل الارتباط أنه مقياس إحصائي يستخدم لبيان نوع العلاقة بين المتغيرات سواء كانت هذه العلاقة طرديه أو عكسية .

ويأخذ معامل الارتباط قيمة من القيم المحصورة بين 1+ و 1- فمعامل ارتباط (1+) بين متغيرين يعبر عن علاقة موجبة تامة (طردية) بين المتغيرين، أما معامل ارتباط (1-) فيعبر عن علاقة تامة ولكنها سالبة(عكسية) بين المتغيرين. ومعاملات الارتباط بين أزواج من المتغيرات والتي تتراوح قيمها بين (1+ و1-) تعبر عن علاقة غير تامة فمعامل ارتباط (0.90 ±) يعبر عن علاقة قوية بين المتغيرات، ومعامل ارتباط (0.50 ±) يعبر عن علاقة متوسطة، أما معامل ارتباط(0.20 ±) فيعبر عن علاقة ضعيفة بين التغيرين.

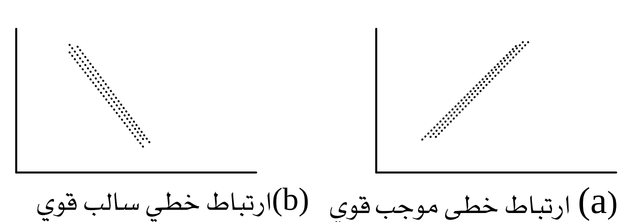
(قيمة معامل الارتباط لا تتغير إذا طرحنا (أو جمعنا) أي عدد ثابت من جميع قيم الظاهرتين).
(قيمة معامل الارتباط لا تتغير إذا ضربنا أو قسمنا جميع قيم الظاهرتين على أي عدد ثابت.

2.2 شكل الانتشار (Scatter Diagram):

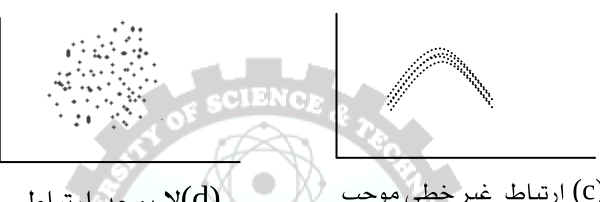
يمكننا تقدير أو تحديد قيمة المتغير التابع من قيمة المتغير المستقل من خلال العلاقة الرياضية بين المتغيرات المستقلة والتابعة، أما إذا أردنا تحديد قوة واتجاه العلاقة بين متغير وآخر فإننا نحتاج لعرض شكل الانتشار لتمثيل العلاقة بين ظاهرتين باستخدام الرسم البياني(للمتغير المستقل X والذي يمثل المحور الأفقي، والمتغير التابع Y والذي يمثل المحور العمودي). وتمثل كل ثنائية (زوج)من القيم (x_i , y_i) بنقطة في المستوى الأحداثي فنحصل على شكل الانتشار والذي يتوقف على نوع العلاقة بين المتغيرين والشكل التالي رقم (1.3) يعرض بعض الأشكال المختلفة للانتشار والتي نلاحظ فيها وجود الارتباط الخطي من عدمه، أي تحديد اتجاه وقوة العلاقة بين المتغيرين.



شكل(3 – 1)



(a) ارتباط خطي موجب قوي (b)ارتباط خطي سالب قوي



(c) ارتباط غير خطي موجب قوي (علاقة غير خطية) (d)لا يوجد ارتباط

من خلال الأشكال السابقة ومن قراءة هذه الأشكال يمكن تحديد طبيعة ونوع ومتانة العلاقة بين المتغيرات. فإذا كانت نقاط الانتشار تقع في مجال متقارب وعلى امتداد خط مستقيم واحد من أسفل إلى أعلى فأننا نستنتج وجود علاقة خطية طردية قوية بين المتغير x والمتغير y كما في الشكل (a) من شكل رقم (1.3). أما إذا كان شكل الانتشار في حدود خط مستقيم يتجه من أعلى إلى أسفل وبدرجة تركيز عالية حول الخط فإن هذا يدل على وجود علاقة خطية عكسية قوية بين المتغيرين x , y كما في الشكل (b). أما إذا اظهر شكل الانتشار تركيزاً للنقاط ولكن حول منحنى وليس خط مستقيما فإننا نستنتج من ذلك عدم وجود علاقة خطية بين المتغيرين x , y كما في الشكل (c). أما إذا كان شكل الانتشار أكثر تشتتاً للنقط بحيث يتعذر معه أن تقع في شكل خط مستقيم أو منحنى فإن ذلك يدل على عدم وجود علاقة ارتباطيه بين المتغيرين x , y كما في شكل (d).

ونخلص من ذلك إلى أن الأشكال السابقة هي عبارة عن أمثلة لأشكال الانتشار التي تبين مدى قوة أو ضعف الارتباط بين المتغيرين (x , y). أي أنها تعطي فكرة واضحة فيما إذا كانت y تزداد بزيادة قيمة x أو تنقص بزيادة x أو أنها لا تتأثر.

3.2 حساب معامل الارتباط بين متغيرين:

يمكن حساب معامل الارتباط بعدة طرق ونذكر هنا أكثرها شيوعاً:

1- معامل ارتباط بيرسون (Pearson): يستخدم عندما يكون كلا المتغيرين كمي ويقاس بمقياس فئوي مثل أيجاد الارتباط بين الدخل والإنفاق.

2- معامل ارتباط سبيرمان (Spearman): والذي يسمى أيضاً معامل ارتباط الرتب ويستخدم عندما يكون المتغيرين مقاسين بمقياس ترتيبي مثل إيجاد الارتباط بين متغير مستوى الدخل(عالي، متوسط، منخفض) ومتغير تأييد رأي معين (أوافق، متردد، غير موافق) كما يستخدم معامل (سبيرمان) عندما يكون كلا المتغيرين مقاساً بمقياس فئوي.

إشارة معامل الارتباط

أ- إذا كان $r > 0$ فإن الارتباط يكون إيجابي (طردي) .

ب- إذا كان $r < 0$ فإن الارتباط يكون سلبي (عكسي).

ج- إذا كان $r = 0$ فإنه لا يوجد ارتباط.

4.2 قيمة معامل الارتباط:

كلما زادت القيمة المطلقة لمعامل الارتباط كلما اقتربت النقط من الخط المستقيم، ويكون الارتباط خطي قوي. وكلما قلت القيمة المطلقة لمعامل الارتباط كلما ابتعدت النقط عن الخط المستقيم، ويكون الارتباط (ضعيف).

أما إذا كانت $r = \pm 1$ فأن جميع النقط تقع على الخط المستقيم، ويكون الارتباط تام. وبشكل عام توصف معاملات الارتباط التي تزيد عن (0.8) بأنها قوية، والتي تقع حول (0.5) بأنها متوسطة، التي تقل عن (0.3) بأنها ضعيفة.

نظرية الارتباط:

نفرض أن $x_1, x_2, ..., x_n$ تمثل قياسات ظاهرة ما ولتكن X . ونفرض أن $y_1 , y_2, ..., y_n$ تمثل قياس ظاهرة أخرى ولتكن Y . ويعرف معامل الارتباط r بين الظاهرتين X و Y على أنه:

$$r=\frac{\sum_{i=1}^{n} X_i\ Y_i}{\sqrt{\left(\sum x_i^2\right)\left(\sum y_i^2\right)}}=\frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{\sqrt{\left(\sum_{i=1}^{n} X_i^2\right)\left(\sum_{i=1}^{n} Y_i^2\right)}}$$

حيث

$$X_i=X_i-\bar{X}$$

$$Y_i=Y_i-\bar{Y}$$

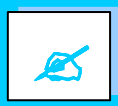
أو باستخدام العلاقة:

$$r=\frac{\sum X_i\ Y_i}{S_x\ S_y}$$

حيث S_x : الانحراف المعياري للمتغير المستقل X

S_y : الانحراف المعياري للمتغير التابع Y

مثال (1.3):



أوجد معامل الارتباط بين X , Y حيث

X :1 2 4 5

Y :2 6 10 10

الحل :

$$\bar{x}=\frac{12}{4}=3\quad ,\quad \bar{y}=\frac{28}{4}=7$$

جدول (1.3) يوضح كيفية حساب معامل الارتباط

i	X	Y	$X_i-\bar{X}$	$Y_i-\bar{Y}$	$X_i\ Y_i$	X_i^2	Y_i^2
1	1	2	-2	-5	10	4	25
2	2	6	-1	-1	1	1	1
3	4	10	1	3	3	1	9
4	5	10	2	3	6	4	9
$\sum$	12	28	--	--	20	10	44

حيث i=1,2,3,4

$$r=\frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{\sqrt{\left(\sum_{i=1}^{n} X_i^2\right)\left(\sum_{i=1}^{n} Y_i^2\right)}}$$

$$\because \quad r=\frac{\sum_{i=1}^{n} X_i \; Y_i}{\sqrt{\left(\sum x_i^2\right)\left(\sum y_i^2\right)}}=\frac{\sum_{i=1}^{n} X_i \; Y_i}{\sqrt{\left(\sum_{i=1}^{n} X_i^2\right)\left(\sum_{i=1}^{n} Y_i^2\right)}}$$

$$\therefore \quad r=\frac{20}{\sqrt{(10)(44)}} \quad = \quad 0.95$$

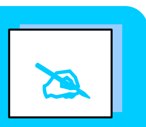
تشير نتيجة التحليل إلى وجود ارتباط خطي طردي قوي بين المتغيرين $(X \;, Y)$ ، أي أنه كلما زادت قيمة المتغير x زادت معها قيمة المتغير Y ، وهذا يؤيد ما تم استنتاجه عند رسم شكل الانتشار.

واضح من المثال السابق أن العمليات الحسابية كانت سهله وذلك لآن كلا من الوسط الحسابي للمتغير X والوسط الحسابي للمتغير Y عدد صحيح ولكن إذا كان أي من المتوسطين عدد كسري (ذو أرقام عشرية) . فإن الحل يكون من الصعوبة بمكان لذلك نلجأ إلى صورة أخرى لمعامل الارتباط:

$$r=\frac{n\sum_{i=1}^{n} X_i Y_i-\left(\sum X_i\right)\left(\sum Y_i\right)}{\sqrt{\left(n\sum X_i^2-\left(\sum X_i\right)^2\right)\left(n\sum Y_i^2-\left(\sum Y_i\right)^2\right)}}$$

أو باستخدام الصيغة:

$$r=\frac{\sum X_i Y_i-n\bar{X}\bar{Y}}{\sqrt{\left(\sum_{i=1}^{n} X_i^2-n\bar{X}\right)\left(\sum_{i=1}^{n} Y_i^2-n\bar{Y}\right)}}$$



مثال (2.3):

باستخدام الصورة الأخيرة لمعامل الارتباط أوجد معامل الارتباط للمثال (1.3) .

جدول (2.3)

i	X	Y	$X_i Y_i$	X^2	Y^2
1	1	2	2	1	4
2	2	6	12	4	36
3	4	10	40	16	100
4	5	10	50	25	100
$\sum$	12	28	104	46	240

الحل :

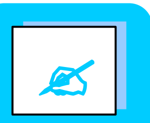
من جدول (2.3)

$$r=\frac{(4)(104)-(12)(28)}{\sqrt{\left(4(46)-(12)^2\right)\left(4(240)-(28)^2\right)}}$$

$$\therefore r=\frac{416-336}{\sqrt{(184-144)(960-784)}}=\frac{80}{\sqrt{(40)(176)}}$$

$$\therefore r=\frac{80}{\sqrt{7040}}=0.95$$

وعلى الدارس محاولة حل المثال بالطريقة الأخيرة، وينطبق التعليق السابق نفسه على النتيجة هنا.



مثال(3.3):

البيانات الآتية تمثل نتائج الاختبار النظري (X) والاختبار العملي (Y) لستة أشخاص تقدموا للتوظيف في إحدى الشركات.

المطلوب: احسب معامل الارتباط وفسر النتيجة التي حصلت عليها.

X_i	31	40	25	30	20	34
Y_i	5	11	3	4	2	5

الحل

نكون الجدول التالي رقم(3.3)

جدول(3.3)

i	X	Y_i	X_iY_i	X_i^2	Y_i^2
1	31	5	155	961	25
2	40	11	440	1600	121
3	25	3	75	625	9
4	30	4	120	900	16
5	20	2	40	400	4
6	34	5	170	1156	25
$\sum$	180	30	1000	5642	200

باستخدام الصيغة:

$$r=\frac{n\sum_{i=1}^{n}X_i Y_i-(\sum X_i)(\sum Y_i)}{\sqrt{\left(n\sum X_i^2-(\sum X_i)^2\right)\left(n\sum Y_i^2-(\sum Y_i)^2\right)}}$$

نجد أن:

$$r=\frac{6(1000)-(180)\,(30)}{\sqrt{[\,6(5642)-(180)^2\,]\,[6(200)-(30)^2}}$$

$$r\;=\frac{600}{\sqrt{(1452)\,(300)}}=0.91$$

تبين النتيجة أن العلاقة بين المتغيرين $(x\,,\,y)$ خطية طردية لأن الإشارة موجبة كما أنه ارتباط قوي.ويمكن حلها باستخدام الصيغة:

$$r=\frac{\sum X_i Y_i-n\bar{X}\,\bar{Y}}{\sqrt{\left(\sum_{i=1}^{n}X_i^2-n\bar{X}\right)\left(\sum_{i=1}^{n}Y_i^2-n\bar{Y}\right)}}$$

فنلاحظ أن:

$$\bar{x}=\frac{\sum x_i}{n}=\frac{180}{6}=30$$

$$\bar{y}=\frac{\sum y_i}{n}=\frac{30}{6}=5$$

$$\therefore\; r=\frac{1000-(180)\,(30)}{\sqrt{[\,5642-6\,(30)^2\,]\,[\,200-6(5)^2}}$$

$$r\;=\frac{100}{\sqrt{(242)\,(50)}}=0.91$$

وهي النتيجة السابقة نفسها عند تطبيق الصيغة الأولى.

UNIVERSITY OF SCIENCE & TECHNOLOGY
جامعة العلوم والتكنولوجيا

5.2 خصائص معامل ارتباط بيرسون

1-تتراوح قيمة معامل الارتباط r بين (± 1)

2- تكون قيمة معامل الارتباط r (± 1) عندما تكون العلاقة بين المتغيرين تامة وعندما لا تساوي أي منها تكون غير تامة.

3- قيمة معامل الارتباط تساوي الصفر عندما لا توجد علاقة بين المتغيرين.

4- تكون العلاقة طردية عندما يكون معامل الارتباط موجب وتكون العلاقة عكسية عندما يكون معامل الارتباط سالب.

من أهم مزايا معامل الارتباط أن قيمته مجردة من وحدات القياس.

ومن أهم مساوئ معامل الارتباط لبيرسون هي أنه لا يعبر عن متانة العلاقة بشكل صحيح إلا إذا كانت خطية، أما إذا كانت العلاقة عبارة عن منحنى فقيمته لا تعبر عن متانة العلاقة. لذا يجب في مثل هذه الحالات استخدام مقاييس أخرى كما أنه لا يمكن استخدامه إذا كانت المتغيرات نوعية(وصفية).



عرف الارتباط. ومعامل الارتباط مع ذكر بعض الأمثلة على العلاقات ألارتباطيه.



احسب معامل الارتباط بين المتغيرين (x , y) حيث:

X_i : 6 4 8 5 7

Y_i : 2 6 5 7 5



اذكر أهم خصائص معامل الارتباط، وأهم مميزاته، وعيوبه.





أسئلة التقويم الذاتي (1) :

1) إذا كان معامل الارتباط بين متغيرين يساوي (0.95) . فما قوة واتجاه العلاقة بينهما؟

2) ماذا نقصد بشكل الانتشار؟

3) أوجد معامل الارتباط بين X , Y حيث :

X_i	10	7	6	2 :
Y_i	11	8	5	3 :

6.2 معامل ارتباط الرتب(سبيرمان) (Rank Correlation Coefficient)

معامل الارتباط الخطي البسيط والذي سبق الحديث عنه يقيس مقدار الارتباط بين متغيرين وذلك في حالة البيانات الكمية. أما إذا كانت البيانات وصفية فإننا نحتاج في حالات كثيرة إلى إيجاد معامل الارتباط ليس بين القيم الأصلية وإنما بين الرتب التي تأخذها هذه القيم، أي تكون الظواهر غير قابلة للقياس أو التي لا يمكن قياسها لأسباب قاهرة مثل تقديرات الطلاب في مادتين مختلفتين أو تحديد كمية الغيوم المتواجدة في سماء منطقة ما، لذلك نقول اصطلاحاً أن هناك غيوم كثيفة أو متوسطة أو قليلة أو أن السماء صافية ،وبهذا نكون قد عبرنا عن كمية الهطول أو كثافتها بدون أن نقيسها ، فتكون التقديرات هي A , B , C , D , E ,.... فإنه يصعب حساب معامل الارتباط بطريقة (بيرسون)، لذلك نشأت الحاجة إلى مقياس يعطي قوة الارتباط للبيانات الوصفية . وهذا المقياس يسمى بمعامل ارتباط الرتب لسبيرمان وهو مقياس للارتباط ولكلا البيانات الكمية و الوصفية التي لها صفة الترتيب مثل تقديرات الطلاب أو وصف حالة الجو ....الخ، حيث يمكن إعطاء رتب لها من حيث كبر وصغر التقدير. وحساب معامل ارتباط الرتب يقترب كثيراً من معامل ارتباط (بيرسون) ولكن يمتاز عنه بسهولة حسابه وخاصة عندما يكون أزواج القيم أقل من $(n<30)$.

ويحسب معامل ارتباط الرتب من العلاقة:

$$r_{rank} = 1 - \frac{6\sum D^2}{n(n^2-1)}$$

حيث

D = الفروق بين رتب القيم المتقابلة في Y, X

N = عـدد أزواج القـيم (Y, X) في البيانـات وتسـمى هـذه الصـيغة معامـل سبيرمان لارتباط الرتب.

ويتم حساب معامل ارتباط الرتب باستبدال الصفات النوعية بقيم عددية تعبر عنها تدعى الرتب ونبدأ بأضعف المشاهدات ثم الأقوى أو العكس.

والمقصود بالرتب هنا إيجاد رتب القراءات Y, X مع بقاء كل قراءة مكانها وذلك بأن نتصور ترتيب البيانات تصاعدياً أو تنازلياً، ففي حالة الترتيب التصاعدي تكون أصغر قراءة قيمتها تساوي واحد والقيمة التي تليها تأخذ الرتبة 2 وهكذا.

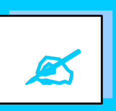
وفي حالة تساوي قيمتين أو أكثر من القيم المختلفة فإن هناك رأيين:

1- إعطاء هذه القيم نفس الرتبة وإهمال ما كان سيعطى لها من ترتيب فيما لو كانتا غير متساويتين.

2- إعطاء كل قيمة من القيم المتساوية ترتيباً يساوي الوسط الحسابي لترتيب القيم المتكررة.

ويستخدم معامل ارتباط الرتب مثله مثل معامل بيرسون لقياس متانة العلاقة بين متغيرين وتحديد نوع العلاقة أيضاً فضلاً عن معرفة وجود العلاقة بين المتغيرين من عدمها.

مثال(4.3):



أوجد رتب المتغير X التي قيمها معطاة كما يلي:

95 ,81, 72 , 86 , 75 , 86

الحل:

حسب الرأي الأول نقوم بترتيب القراءات كما يلي:

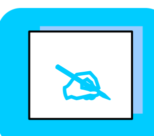
قيم X	95	86	86	81	75	72
رتبة X	6	4	4	3	2	1

نقوم بإعطاء كل قيمة من القيم المتساوية ترتيباً يساوي الوسط الحسابي لترتيبهما فيصبح الترتيب على الصورة:

قيم X	95	86	86	81	75	72
رتبة X	6	4.5	4.5	3	2	1

ولما كان الرأي الثاني يتضمن رتب كسرية ينتج عنها صعوبة في العمليات الحسابية فإن البعض يفضل إتباع الرأي الأول تسهيلاً للحساب.إلا أن الشائع هو استخدام الرأي الثاني لأنه يأخذ جميع الرتب في الاعتبار ولا يهمل أي رتبة.

مثال (5.3):



الجدول الآتي يوضح درجات مجموعة مكونة من ثمانية طلاب في كل من مادتي الإحصاء والمحاسبة في أحد الامتحانات النهائية.

الإحصاء X_i	13	9	19	15	11	8	16	11
المحاسبة Y_i	15	7	17	15	10	9	14	10

المطلوب:

1- هل هناك علاقة بين تحصيل الطلاب في المادتين ؟ وإذا وجدت علاقة حدد نوعها وقوتها.

2- أوجد معامل ارتباط الرتب لدرجات الطلاب لمادتي الإحصاء والمحاسبة.

الحل :

1- نوجد معامل ارتباط بيرسون

جدول(4.3)

i	X_i	Y_i	X_iY_i	X_i^2	Y_i^2
1	13	15	195	169	225
2	9	7	63	81	49
3	19	17	323	361	289
4	15	15	225	225	225
5	11	10	110	121	100
6	8	9	72	64	81
7	16	14	224	256	196
8	11	10	110	121	100
$\sum_{i=1}^{n}$	102	97	1322	1398	1265

الحل:

$$r = \frac{8(1322) - (102)(97)}{\sqrt{\{8(1398) - (102)^2\}\{8(1265) - (97)^2\}}}$$

$$= \frac{10576 - 9894}{\sqrt{(780)(711)}} = \frac{682}{\sqrt{554580}}$$

$$= \frac{682}{744.701} = 0.92$$

نعم توجد علاقة بين تحصيل الطلاب في المادتين . وهي علاقة طردية قوية جداً

2- لحساب معامل ارتباط الرتب نكون الجدول(5.3) الآتي:

جدول(5.3)

X_i	Y_i	رتب X_i	رتب Y_i	D	D^2
13	15	5	6.5	-1.5	2.25
9	7	2	1	1	1
19	17	8	8	0	0
15	15	6	6.5	-0.5	0.25
11	10	3.5	3.5	0	0
8	9	1	2	-1	1
16	14	7	5	2	4
11	10	3.5	3.5	0	0
					$\sum D^2 = 6.5$

$$r_{rank} = 1 - \frac{6\sum D^2}{N(N^2-1)}$$

$$= 1 - \frac{51}{504} = 0.90$$

جامعة العلوم والتكنولوجيا

وهذا معناه أن العلاقة بين رتب الدرجات النهائية لمقرري الإحصاء والمحاسبة طردية وقوية جداً وهي نتيجة مقاربة للنتيجة التي توصلنا إليها عند حساب معامل الارتباط الخطي في الفقرة (1) والذي كان يساوي (0.92) وهي قيمة قريبة. وينبغي عليك عزيزي الدارس الانتباه إلى أن معامل ارتباط الرتب (سبيرمان) هو للارتباط بين الرتب وليس للقيم الأصلية ، بينما معامل (بيرسون) هو للارتباط بين القيم الأصلية ومما لا شك فيه أن معامل ارتباط (بيرسون) هو الأكثر كفاءة ودقة في حالة كون قيم المتغيرين كمية كما في مثالنا هذا.

مثال(6.3):

لنفترض أنه لدينا المعلومات التالية عن متوسط عدد الأطفال في الأسرة والمستوى التعليمي للأم :

المستوى التعليمي للأم	أمية	تقرأ وتكتب	ابتدائية	إعدادية	ثانوية	جامعية	دكتوراه
متوسط عدد الأطفال	10	8	9	7	5	4	2

المطلوب: تقدير متانة ونوع العلاقة الارتباطية بين المستوى التعليمي للأم ومتوسط عدد الأطفال في الأسرة ؟

الحل :

نعطي للمتغيرين رتب تتناسب مع كل صفة كما يوضحها الجدول رقم (6.3)

جدول(6.3)

i	المستوى التعليمي للأم	متوسط عدد الأطفال	رتب المستوى التعليمي للأم	رتب متوسط عدد الأطفال	D	D^2
1	إعدادية	7	4	4	0	0
2	ابتدائية	9	3	6	-3	9
3	ثانوية	5	5	3	2	4
4	أميه	10	1	7	-6	36
5	تقرأ وتكتب	8	2	5	-3	9
6	جامعية	4	6	2	4	16
7	دكتوراه	2	7	1	6	36
$\sum$	--	---	--	--	--	110

$$\because \quad r_{rank} = 1 - \frac{6\sum D^2}{N\,(N^2-1)}$$

$$\therefore \quad r_{rank} = 1 - \frac{6\,(110)}{7\,(49-1)}$$

$$r_{rank} = 1 - 1.964 = -\,0.96$$

تشير النتيجة وجود علاقة ارتباط متينة بين المستوى التعليمي للأم ومتوسط عدد الأطفال في الأسرة وهي علاقة عكسية قوية جداً، وهذا يدلل على أنه كلما زاد المستوى التعليمي للأم كلما قل متوسط عدد الأطفال في الأسرة.



تدريب (4)

الجدول التالي يوضح الترتيب الأبجدي لعشرة طلاب حسب مستوى أدائهم في كل من الجانب العملي والجانب النظري لمادة البيولوجي أوجد معامل ارتباط الرتب

الجانب العملي	5	1	6	4	10	7	2	9	3	8
الجانب النظري	6	2	4	3	7	8	1	10	5	9

أسئلة التقويم الذاتي (2) :

مع أيٍّ من المتغيرات يفضل استخدام معامل ارتباط الرتب لسبيرمان

7.2 ارتباط الصفات (Correlation of Attributes)

أوضحنا سابقاً بأن معامل ارتباط (بيرسون) يقيس قوة الارتباط في حالة البيانات الكمية سواء كانت مبوبة أو غير مبوبة، وكذلك معامل ارتباط الرتب (لسبيرمان) والذي يستخدم لحساب قوة الارتباط في حالة البيانات الكمية والوصفية التي لها صفة الترتيب إلا أنه قد توجد بيانات وصفية لها صفات مميزة ولا نستطيع ترتيبها مثل الحالة الاجتماعية (متزوج، أعزب، مطلق، أرمل) وكذلك لون العينين (زرقاء، سوداء، بنية) ولقياس العلاقة (قوة الارتباط) لهذه البيانات نشأت الحاجة إلى إيجاد مقاييس تقيس الارتباط بين هذه الصفات وسنذكر منها:

1.7.2 معامل الاقتران (Coefficient of Association):

يستخدم لقياس قوة الارتباط عندما يكون لدينا متغيرين نوعيين ولكل متغير صفتين فقط أي يكون لدينا جدول اقتران 2 * 2 ، كأن تدرس العلاقة بين الإصابة بالسرطان والتدخين لمجموعة أشخاص مدخنين وغير مدخنين فالمتغيرين هنا الأول المرض (مصاب، غير مصاب) والثاني التدخين (مدخن، غير مدخن) أو ندرس قياس قوة العلاقة بين التدخين والتعليم فإننا نجد أن التعليم ينقسم إلى (متعلم، غير متعلم) والتدخين (مدخن، غير مدخن) ولا يستخدم معامل الاقتران عندما يكون كل من المتغيرين أو أحدهما له أكثر من صفتين. ومعامل الاقتران كمعامل الارتباط قيمته تكون محصورة بين الصفر والواحد الصحيح، ولحساب معامل الاقتران تقسم المفردات حسب المتغيرات والصفات كما في الجدول (7.3) الآتي:

جدول (7.3)

المتغير A / المتغير B	القيمة (1)	القيمة (2)	المجموع
القيمة (1)	n_{11}	n_{12}	n_1
القيمة (2)	n_{21}	n_{22}	n_2
المجموع	n_1	n_2	N

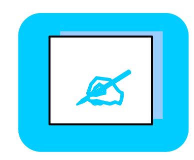
حيث $N = n_{11} + n_{12} + n_{21} + n_{22}$

ويحسب معامل الاقتران من العلاقة

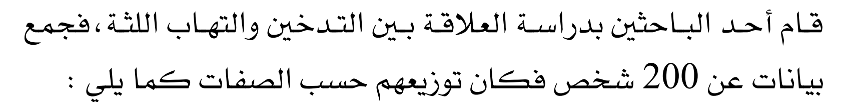
$$r_c = \frac{n_{11} \cdot n_{22} - n_{21} \cdot n_{12}}{n_{11} \cdot n_{22} + n_{21} n_{12}}$$

وتكون هناك علاقة قوية بين المتغيرات كلما اقتربت قيمة معامل الاقتران من الواحد وبشكل عام نقول أن هناك علاقة عندما تكون قيمة r_c أكبر من 0.5

جامعة العلوم والتكنولوجيا UNIVERSITY OF SCIENCE & TECHNOLOGY



مثال (7.3):



قام أحد الباحثين بدراسة العلاقة بين التدخين والتهاب اللثة، فجمع بيانات عن 200 شخص فكان توزيعهم حسب الصفات كما يلي :

B \ A	مدخن	غير مدخن	مجموع
مصاب	70	10	80
غير مصاب	20	100	120
مجموع	90	110	200

أحسب معامل الاقتران ؟

الحل :

$$r_c = \frac{n_{11} \cdot n_{22} - n_{21} \cdot n_{12}}{n_{11} \cdot n_{22} + n_{21} n_{12}}$$

$$r_c = \frac{70 \cdot 100 - 20 \cdot 10}{70 \cdot 100 + 20 \cdot 10} = \frac{6800}{7200} = 0.944$$

وهذا يدل على أن هناك علاقة قوية جداً بين التدخين ومرض التهاب اللثة وتجدر الإشارة إلى أنه ليس هناك أي مدلول لإشارة معامل الاقتران، وذلك لأن إشارته تختلف حسب ترتيب تكرار الصفات في الجدول لذلك تؤخذ قيمته المطلقة للدلالة على وجود العلاقة من عدمها. وقيمة هذا المعامل تكون دائماً أقل من الواحد الصحيح ما دامت $n_{21}.n_{12} > 0$ ، ولا تساوي الواحد إلا إذا كانت $n_{21}.n_{12} = 0$

2.7.2 معامل التوافق (Coefficient of Contingency) :

نعلم من دراستنا السابقة أن معامل الاقتران يستخدم لقياس متانة العلاقة بين متغيرين نوعيين لكل منهما قيمتين (أن الجدول يحتوي على أربع خانات) ولا نستطيع استخدامه إذا كان لأحد المتغيرين على الأقل أكثر من صفتين. وفي هذه الحالة يكون الجدول مكون من أكثر من أربع خلايا فنلجأ إلى استخدام مقياس آخر هو معامل التوافق ويرمز له بالرمز r_c ولحساب معامل التوافق نفرض أن لدينا ظاهرة X لها r من الصفات والظاهرة Y لها s من الصفات ونوضح ذلك في الجدول (8.3) :

جدول(8.3)

X \ Y	x_1	x_2	....	x_r	المجموع
y_1	f_{11}	f_{12}	....	f_{1r}	$f_{1.}$
y_2	f_{21}	f_{22}	....	f_{2r}	$f_{2.}$
. .	. .	. .	. .	. .	. .
y_s	f_{s1}	f_{s2}	....	f_{rs}	$f_{s.}$
المجموع	$.f_{.1}$	$f_{.2}$	....	$f_{.r}$	$f_{..}$

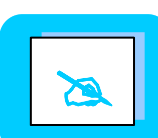
ويحسب معامل التوافق r_c بالعلاقة الآتية:

$$r_c = \sqrt{\frac{G-1}{G}}$$

حيث :

$$G=\frac{(f_{11})^2}{f_{.1}\times f_{1.}}+\frac{(f_{12})^2}{f_{.2}\times f_{1.}}+....+\frac{(f_{rs})^2}{f_s\times f_{r.}}$$

مثال (8.3):



من بيانات الجدول التالي أوجد معامل التوافق بين متغيري الحالة الاجتماعية والتدخين .

الحالة الاجتماعية \ التدخين	a	b	c	المجموع
يدخن	30	80	20	130
لا يدخن	25	15	30	70
المجموع	55	95	50	200

الحل : لدينا :

$$G=\frac{(30)^2}{(55)(130)}+\frac{(80)^2}{(95)(130)}+\frac{(20)^2}{(50)(130)}+\frac{(25)^2}{(55)(70)}+\frac{(15)^2}{(95)(70)}+\frac{(30)^2}{(50)\,(70)}$$

$$= 1.15946$$

وبتطبيق الصيغة

$$r_c = \sqrt{\frac{G-1}{G}} = \sqrt{\frac{0.15946}{1.15946}} = 0.37$$

التي تشير إلى علاقة متوسطة القوة بين التدخين والحالة الاجتماعية للشخص مع التنويه إلى أن زيادة عدد الأعمدة والصفوف من شأنها أن تزيد في ارتفاع معامل التوافق عموماً، لذلك فإن أصغر قيمة لـ r_c هي الصفر وأكبر قيمة تعتمد على عدد الصفوف و عدد الأعمدة ولكنها في كل الأحوال لا تتجاوز الواحد .

تدريب (5)



لنفترض أننا نريد التأكد من وجود أو عدم وجود علاقة ارتباطيه بين الوضع النهائي للطبة (ناجح ،أو راسب) وجنس كل منهم (ذكر ، أو أنثى) فأخذنا عينة من 100 طالب فكانت النتائج حسب الجدول

وضع الطالب / الجنس	ناجح	راسب	مجموع
ذكر	65	5	70
أنثى	27	3	30
مجموع	92	8	100

أوجد معامل التوافق ؟

أسئلة التقويم الذاتي (3) :

1) مع أي من البيانات يستخدم معامل الاقتران.

2) ما وجه الشبه بين معامل الاقتران ومعامل التوافق.

3. الانحدار (Regression):

عزيزي الدارس، في الجزء السابق من هذه الوحدة تعرضنا لمفهوم الارتباط والارتباط الخطي وتعرفت على كيفية حساب قيمة معامل الارتباط الخطي بين متغيرين كميين أو وصفيين كما تعرفت على كيفية قياس قوة العلاقة الخطية بينهما واتجاه هذه العلاقة سواءاً كانت طردية أم عكسية إضافة إلى دراستك لشكل الانتشار والذي من خلاله يكون التشخيص الأولي للعلاقة بين المتغيرين.

وفي هذا البند سوف نتناول مفهوم الانحدار وأنواعه وأهم استخداماته وتطبيقاته حيث إن "تحليل الانحدار" يعتبر من أكثر الطرق الإحصائية استخداماً في مختلف العلوم لأنه يصف العلاقة بين المتغيرات على هيئة معادلة، وأسلوب مهم من أساليب الإحصاء التطبيقي في التحليل الإحصائي كونه الأداة الفاعلة لدراسة العلاقة بين المتغيرات. وقد توسعت ولازالت تتوسع تطبيقاته حتى شملت ليس فحسب العلوم التطبيقية بل حتى العلوم الإنسانية. ولعل من أهم تطبيقاته تلك التي تتعلق بمسألة التقدير أو التخمين أو التنبؤ.

إن دراسة الانحدار هي عبارة عن تمثيل العلاقة بين المتغيرات بمعادلة رياضية ومن ثم استنتاج متغيرات العلاقة اعتماداً على هذه المعادلة الرياضية.

ويمكن تقسيم الانحدار إلى الأنواع التالية وذلك حسب عدد المتغيرات المستقلة الداخلة بالمعادلة وحسب نوع معادلة التمثيل:

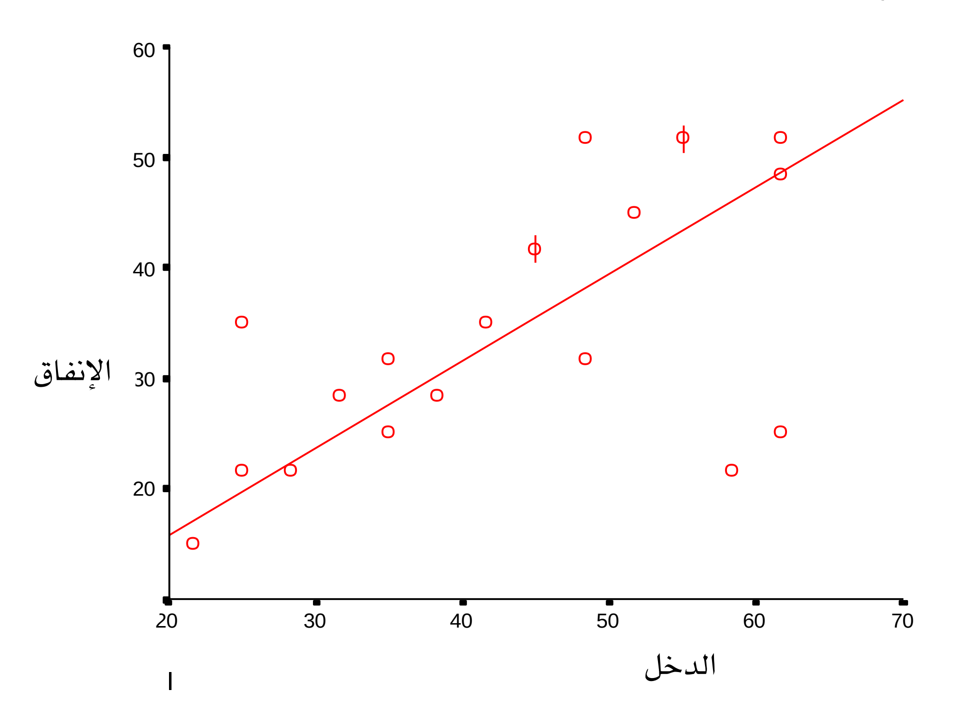
- الانحدار الخطي البسيط: وذلك عندما يمكن تمثيل العلاقة بخط مستقيم.
- الانحدار غير الخطي البسيط : وذلك عندما تكون العلاقة بين المتغيرات يمكن تمثيلها بعلاقة غير معادلة الخط المستقيم:(درجة ثانية أو ثالثة أو أسية أو لوغاريتمية).
- الانحدار المتعدد : وهو دراسة العلاقة بين عدة متغيرات مستقلة ومتغير واحد تابع ويقسم إلى قسمين:
- الانحدار الخطي المتعدد: وذلك عندما يمكن تمثيل العلاقة بين المتغيرات المستقلة والمتغير التابع بمعادلة خطية.
- الانحدار المتعدد اللاخطي:وتكون بتمثيل العلاقة بين المتغيرات المستقلة والمتغير التابع بعلاقة غير خطية.

تتميز طريقة المربعات الصغرى بأنها تعطينا خطأً يكون مجموع مربعات انحرافات النقط عنه أصغر ما يمكن.

1.3 شكل الانتشار (Scatter Diagram)

يمكن استعمال الرسم البياني (شكل الانتشار) والذي يقصد به تمثيل العلاقة بين متغيري ظاهرتين على المحاور الإحداثية بحيث يكون المحور الأفقي ممثلاً للمتغير المستقل X والمحور العمودي ممثلاً للمتغير التابع Y ويتم تمثيل أزواج كل من المتغيرين بنقطة واحدة (Y, X) على المستوى الإحداثي فنحصل على شكل الانتشار الذي يتوقف مظهره على نوع العلاقة بين المتغيرين.

شكل (2.3) يوضح شكل الانتشار وكذلك خط الانحدار بين متغيري الإنفاق والدخل



2.3 الانحدار الخطي البسيط

تحليل الانحدار عبارة عن وسيلة إحصائية تستخدم لتحديد نوع العلاقة بين متغير مستقل واحد أو أكثر ومتغير تابع واحد، وعند وجود علاقة خطية بسيطة بين متغير تابع Y ومتغير مستقل X يستخدم عادة نموذج الانحدار الخطي الآتي:

$$Y_i = \propto + \beta X + e_i$$

حيث

Y : المتغير التابع أو الاستجابة في التجربة .

$\propto$: معامل ثابت أو الجزء المقطوع من محور Y ويصبح مساوياً لقيمة Y عندما $x = 0$

β : يدعى بميل الانحدار ويمثل مقدار التغير في Y عند زيادة المتغير X بمقدار وحده واحدة، كما يطلق عليه بمعامل الانحدار.

X_i : المتغير المستقل في التجربة .

e : الخطأ العشوائي : وهو متغير عشوائي يفترض أنه يتبع التوزيع الطبيعي بمتوسط $E(\varepsilon_i)=0$ وتباين $\text{var}(\varepsilon_i)=\sigma^2$ أي أن: $\varepsilon_i \sim N(0,\sigma^2)$ والتباين المشترك بين $\varepsilon_i , \varepsilon_j$ يساوي صفر أي أن: $Cov(\varepsilon_i,\varepsilon_j)=0$ لكل $i, j : i \neq j$.

وقد تم إدراج حد الخطأ العشوائي e في نموذج الانحدار لأسباب منها:

1. يوجد عنصر أساسي لا يمكن التنبؤ به هو العشوائية في الاستجابات البشرية التي يمكن التعبير عنها ضمن حد الخطأ.
2. الصياغة الناقصة حال بناء النموذج الرياضي.
3. لتلافي وجود أخطاء المشاهدات أو القياس.

ويسمى النموذج السابق بنموذج الانحدار الخطي البسيط، كما أن معادلة الانحدار تدلنا على القيمة المتوقعة لـ Y لأنها الأداة التي يمكن من خلالها الحصول على تقدير للمتغير Y ، ومن الناحية العملية لن نحصل على معادلة الانحدار وإنما تقديراً لها من عينة من البيانات الإحصائية باستخدام طريقة المربعات الصغرى.

ولتقدير معلمات الانحدار في النموذج نستخدم طريقة المربعات الصغرى وهي من أكثر الطرق استخداماً في تقدير معلمات الانحدار للمجتمع وتتلخص في جعل

مجموع مربعات الأخطاء العشوائية أقل ما يمكن وتقدر $\propto, \beta$ بالمقدرات $\hat{\propto}$, $\hat{\beta}$ وبعد حسابهما والتعويض في النموذج السابق نحصل على المعادلة التقديرية $\hat{Y} = \hat{\propto} + \hat{\beta} X + e$ حيث $\hat{Y}$ (تقرأ Y هات) تمثل القيمة المقدرة المتنبئة للقيمة Y التي تقابل القيمة X.

للحصول على خط انحدار المتغير التابع Y على المتغير المستقل X بطريقة المربعات الصغرى، فإنه يمكن اعتبار هذا الخط تقديراً لخط انحدار المجتمع $Y = \propto + \beta X$. وهي إحدى الطرق الرياضية لحساب الميل والحد الثابت نفترض أن تكون معادلة الخط المستقيم هي:

$$Y = \propto + \beta X$$

وبأخذ المجموع للطرفين نحصل على

$$\sum Y_i = \sum (\propto + \beta X)$$
$$\sum Y_i = n \propto + \beta \sum X_i$$

وبضرب معادلة الخط المستقيم بالمتغير المستقل X وأخذ المجموع للطرفين نحصل على:

$$\sum X_i Y_i = \propto \sum X_i + \beta \sum X_i^2$$

وبحل هاتين المعادلتين نحصل على:

$$\hat{\beta} = \frac{n\sum X_i Y_i - \sum X_i \sum Y_i}{n\sum X_i^2 - (\sum X_i)^2}$$

وهذه المعادلة يمكن أن تصاغ بصورة أخرى :

$$\hat{\beta} = \frac{\sum X_i Y_i - n\,\overline{X}\overline{Y}}{\sum X_i^2 - n\,\overline{X}^2}$$

وبعد حصولنا على قيمة الميل نستخرج قيمة الحد الثابت وفق العلاقة الآتي:

$$\hat{\propto} = \overline{Y} - \hat{\beta}\,\overline{X}$$

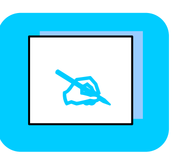
حيث:

$\overline{X}$: تمثل الوسط الحسابي لمشاهدات المتغير المستقل X_i

$\overline{Y}$: تمثل الوسط الحسابي لمشاهدات المتغير التابع Y_i

وبعد تقدير معالم النموذج β ,$\propto$ (قيمة الميل والمقدار الثابت) نعوض بقيمة كل منهما في معادلة الانحدار التقديرية أو دالة الانحدار $\hat{Y} = \hat{\propto} + \hat{\beta} X + e$ فنحصل على معادلة خط انحدار المتغير التابع Y على المتغير المستقل X

مثال(9.3):



احسب معادلة خط انحدار Y على X ثم احسب قيمة y عندما X =10 إذا كانت

X	2	8	4	6	5
Y	3	7	10	4	6

الحل: نوجد الوسط الحسابي لكل من y و X

$$\overline{X} = \frac{\sum X_i}{n} = \frac{25}{5} = 5$$

$$\overline{Y} = \frac{\sum Y_i}{n} = \frac{30}{5} = 6$$

وأيضاً نكون الجدول الآتي:

i	X_i	Y_i	X_iY_i	X_i^2
1	2	3	6	4
2	8	7	56	64
3	4	10	40	16
4	6	4	24	36
5	5	6	30	25
$\sum$	25	30	156	145

نوجد قيمة المقدر $\hat{\beta}$ من العلاقة الآتية:

$$\because \hat{\beta} = \frac{n\sum X_i Y_i - \sum X_i \sum Y_i}{n\sum X_i^{\ 2} - \left(\sum X_i\right)^2}$$

وبالتعويض من الجدول في المعادلة السابقة نحصل على ميل الانحدار $\hat{\beta}$

$$\hat{\beta}_i = \frac{5(156) - (25)(30)}{5(145) - \ (25)^2}$$

$$\hat{\beta} = \frac{780 - 750}{100} = 0.3$$

وبالتعويض أيضاً بقيمة $\hat{\beta}$ في معادلة الحد الثابت

$$\hat{\alpha} = \bar{Y} - \hat{\beta}\ \bar{X}$$

$$\hat{\alpha} = 6 - (0.3) * 5$$

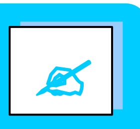
$$\hat{\alpha} = 6 - 1.5 = 4.5$$

لاحظ أن إشارة الحد الثابت موجبة وهذا يطابق ما استنتجناه سابقاً عند تناولنا شكل الانتشار وبذلك تكون معادلة خط انحدار y على x بشكلها النهائي على الصورة:

$$\hat{Y} = 0.3x + 4.5$$

وبالتعويض عندما $X = 10$ في معادلة خط انحدار Y على X نحصل على قيمة $\hat{Y}$ أي:

$$\hat{Y} = 4.5 + (0.3)\ (10) = 7.5$$





مثال(10.3):

لبيانات المثال (5.3) والواردة في جدول(4.3) والتي تبين درجات ثمانية من الطلاب في الإحصاء والمحاسبة أوجد:

1- معادلة خط انحدار درجة المحاسبة y على درجة الإحصاء x

الحل:

من بيانات الجدول(4.3) نلاحظ أن:

$$\sum X_i = 102\ ,\ \sum Y_i = 97\ ,\ \sum X_i\ Y_i = 1322\ ,\ \sum X_i^{\ 2} = 1398\ ,\ \sum Y_i^{\ 2} = 1265$$

وبالتعويض في معادلة الميل $\hat{\beta}$:

$$\because \hat{\beta} = \frac{n\sum X_i Y_i - \sum X_i \sum Y_i}{n\sum X_i^{\ 2} - \left(\sum X_i\right)^2}$$

$$\because\ \hat{\beta} = \frac{8(1322) - (120)\,(97)}{8(1398)\ \ - (102)^2}$$

$$\hat{\beta} = \frac{10576\ - 9894}{11184 - 10404} = \frac{682}{780} = 0.874$$

وللحصول على $\hat{\alpha}$ نستخدم الصيغة التالية:

$$\hat{\alpha} = \bar{Y} - \hat{\beta}\,\bar{X}$$

نحسب $\bar{y}$, $\bar{x}$ كما يلي:

$$\bar{X} = \frac{102}{8} = 12.75$$

$$\bar{Y} = \frac{97}{8} = 12.125$$

$$\therefore\ \hat{\alpha} = 12.125 - 0.874\ (12.75)$$

$$\hat{\alpha} = 12.125 - 11.1435 = 0.9815$$

وعليه فإن معادلة الانحدار التقديرية (أو المعادلة التنبؤية) هي:

$$\hat{Y} = 0.9815 + 0.874\ X$$

ويمكن تفسير معالم نموذج الانحدار للمثال السابق حيث $\hat{\beta} = 0.874$ ، بأنه إذا تحسن مستوى الطالب بمقدار درجة واحدة في مقرر الرياضيات، فإن هذا يؤدي إلى زيادة معدل الطالب في مقرر المحاسبة بمقدار0.874

تدريب (6)



لدراسة العلاقة بين تكاليف الدعاية (X) لمنتج معين وكمية المبيعات (Y) تم الحصول على النتائج التالية:

$$\sum X_i = 81\ ,\ \sum Y_i = 1368\ \ ,\ n = 9$$

$$\sum X_i Y_i = 12668\ ,\ \sum X_i^{\,2} = 789$$

أوجد معادلة خط انحدار Y على X

أسئلة التقويم الذاتي (4) :

- أكمل الفارغات الآتية :

أ- يأخذ معامل الارتباط قيمة محصورة بين ....و..........

ب- يقصد بمعامل الارتباط بأنه مقياس إحصائي يستخدم لبيان .....

ج- يستخدم معامل ارتباط الرتب لقياس ....بين متغيرين

د- من عيوب معامل ارتباط (بيرسون) أنه لا يعبر عن.......بشكل صحيح

هـ- يمكن تقسيم العلاقات بين الظواهر إلى قسمين هما ......و....

و- دراسة الانحدار هي عبارة عن ....

ز- يمكن تقسيم الانحدار حسب النوع إلى ....و......و.......

2-أوجد معادلة خط انحدار Y على X والذي يمثل البيانات التالية

X_i	1	2	4	7	8	9	11
Y_i	1	5	8	10	7	12	11

3- من الجدول الآتي أوجد:

X_i	1	3	4	6	8	9	11	14
Y_i	1	2	4	4	5	7	8	9

1- معامل الارتباط الخطي البسيط للمتغيرين X , Y

2- معادلة خط انحدار Y على X وX على Y

?

4. السلاسل الزمنية (Time Series):

1.4 تعريف السلسلة الزمنية:

تعرف السلسلة الزمنية بأنها: عبارة عن مجموعة القراءات التي تأخذها ظاهرة ما في فترات زمنية متعاقبة غالباً ما تكون متساوية. وتختلف هذه الفترات حسب طبيعة الظاهرة فقد تكون أيام أو أسابيع أو أشهر أو سنين. ويمكن القول بأن السلسلة الزمنية لأي ظاهرة بأنها عبارة عن التطور التاريخي لهذه الظاهرة، فإنتاج الجمهورية اليمنية من البن للأعوام من 1995-2005م أو المبيعات الشهرية لمصنع أسمنت عمران لعام 2004م، أو عدد الزوار اليومي للمتحف الوطني خلال شهر أغسطس لعام 2005م كلها تشكل سلاسل زمنية . أي أن السلسلة الزمنية تحتوي على متغيرين أحدهما يسمى بالمتغير المستقل وهو الزمن (X) والثاني ويسمى بالمتغير التابع وهو قيمة الظاهرة (Y) وبذلك تكون Y دالة في X وتكتب رياضياً:

$$Y = f(x).$$

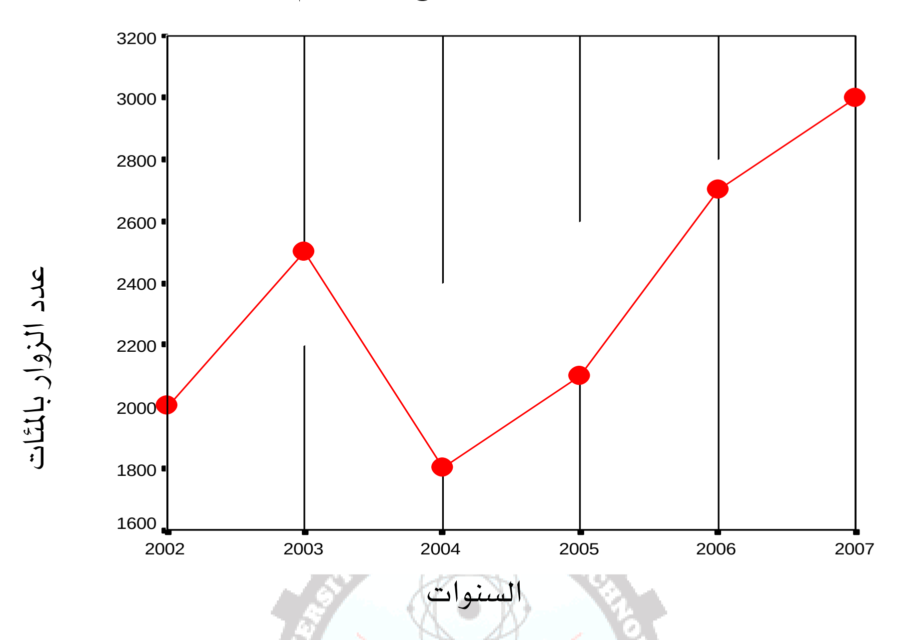
والهدف من الدراسة الإحصائية للسلاسل الزمنية هو الكشف عن التغيرات التي تطرأ على الظواهر التي ندرسها أثناء مدة زمنية معينة حتى يمكن معرفة أنواع هذه التغيرات وقياس كل نوع منها. ويقوم التحليل الإحصائي لأية سلسلة زمنية على أساس المقارنة بين قيم الظاهرة في فترات متتابعة حتى يمكن الكشف عما يصيبها من نمو أو ضمور وحتى تكون المقارنة صحيحة يجب أن تقاس الظاهرة بنفس الوحدات وبنفس الطريقة في الفترات الزمنية المتتابعة

2.4 العرض البياني للسلسلة الزمنية:

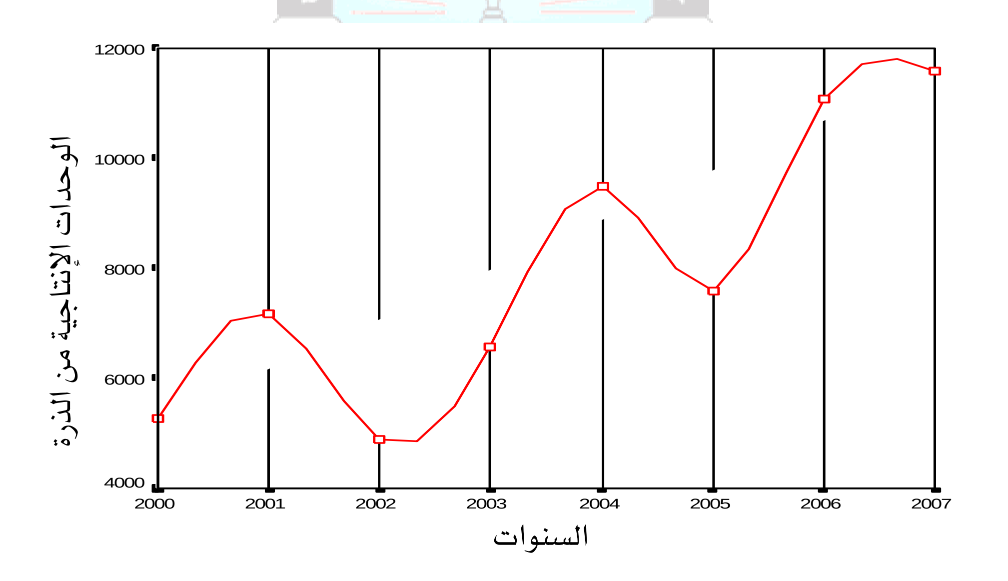
تعرض السلاسل الزمنية بيانياً بأخذ قيم الزمن على المحور الأفقي، وقيم الظاهرة على المحور الرأسي، ثم نضع أمام كل فترة زمنية القيمة المقابلة لها، ثم نصل هذه النقاط فيكون لدينا خط بياني يسمى أحياناً بالمنحنى التاريخي للسلسلة الزمنية أو للظاهرة التي ندرسها.



شكل (3.3) يوضح اتجاه عام موجب



شكل (4.3) يمثل اتجاه عام لسلسلة زمنية



المصدر: بيانات افتراضية

3.4 عناصر السلاسل الزمنية

السلسلة الزمنية: هي عبارة عن سجل تاريخي متتالي يتم اعتماده لبناء التوقعات المستقبلية. وحيث أن قيم أي ظاهرة عبر الزمن تكون تحت تأثير عوامل اقتصادية، واجتماعية، وبيئية، لذا فإن شكل انتشار قيم السلسلة الزمنية عند رسمها على محور أفقي ليمثل السنين ومحور عمودي يمثل قيم هذه السنين سنجد أنها تحت تأثير أربعة أنواع من التغيرات وبدرجات متفاوتة . وهذه المتغيرات يطلق عليها عناصر السلسلة الزمنية. وتتكون السلسلة الزمنية من العناصر الآتية:

1- الاتجاه العام .

2- التغيرات الموسمية.

3- التغيرات الدورية.

4-التغيرات الطارئة (العشوائية).

وسوف نتناول فيما يلي هذه العناصر بشرح موجز لكل من هذه المكونات:

يقصد بتحليل السلسلة الزمنية هو قياس التغيرات التي تؤثر في الظاهرة وخاصة الاتجاه العام والتغيرات الدورية والموسمية لمعرفة مقدار واتجاه وطبيعة كل منها وعزل هذه الأنواع من التغيرات والتنبؤ بقيمة الظاهرة في المستقبل.

1.3.4 الاتجاه العام: Longterm Trend

هو العنصر الذي يقصد به الحركة المنتظمة للسلسلة الزمنية عبر فترة زمنية طويلة نسبياً ويعتبر في العادة أهم عناصر السلسلة الزمنية وغالباً ما يعتمد كعنصر وحيد في بناء التوقعات. ويقال أن الاتجاه العام للسلسلة الزمنية موجباً إذا كان الاتجاه نحو التزايد مع مرور الزمن مثل التزايد في عدد السكان، ويكون سالباً إذا كان الاتجاه نحو التناقص بمرور الزمن مثل نسبة الأميين، إلى مجموع السكان في العديد من دول العالم. وقد يكون الاتجاه موجباً في جزءه الأول وسالباً في جزءه الثاني مثل حالة المبيعات لأجهزة التلفاز غير الملون، وكذلك عدد العمال لبعض الشركات الصناعية التي تقوم لاحقاً باستخدام التكنولوجيا التي تؤدي إلى التقليل من عدد العمال.

وأهم ما يميز الاتجاه العام هو أن التغير الذي يطرأ عليه يكون تدريجي وليس مفاجئ ويظهر أثره واضحاً بعد تراكمه مدة طويلة الأمد حيث تستمر هذه التغيرات في الاتجاه نفسه لفترة طويلة وإذا ما غيرت اتجاهها فإن الاتجاه الجديد يستمر لفترة طويلة أيضاً.

2.3.4 التغيرات الموسمية (Seasonal Variations S) :

هي التغيرات ذات الطبيعة الزمنية الدورية والتي تتكرر بانتظام خلال فترة زمنية لا يزيد طولها على السنة، فقد تكون يومية أو أسبوعية أو شهرية أو فصلية أو سنوية. أي أنها التغيرات المتشابهة التي تظهر في الأسابيع أو الأشهر أو الفصول المتناظرة خلال الفترات الزمنية المختلفة التي تعود إليها مشاهدات السلسلة الزمنية. ومن الأمثلة على ذلك: التغير في عدد المسافرين من ساعة لأخرى أو من يوم لأخر أو من فصل لأخر. وكذلك مبيعات الملابس الشتوية والصيفية وحركة السياحة. وقد تكون أسباب هذه التغيرات الحالة الجوية أو العادات والتقاليد في المجتمع موضوع الدراسة.

3.3.4 التغيرات الدورية (Cyclical Movements C) :

وهي التغيرات التي تطرأ على قيم السلسلة الزمنية بصورة منتظمة أو غير منتظمة وتتكرر في فترات زمنية أكثر من سنة. لذلك فهي تختلف، عن التغيرات الموسمية التي تحدث بانتظام وخلال فترات زمنية أقل من سنة. وتتكون من دوال تشبه دالة الجيب وجيب التمام ولكن بأطوال وسعات قد تكون مختلفة ومن الأمثلة على ذلك الدورات الاقتصادية التي تمر بها بعض الدول حيث يمر الاقتصاد فيها بمرحلة النمو السريع تعقبها مرحلة التراجع الاقتصادي ثم مرحلة الركود ثم استعادة النشاط الاقتصادي ذات النمو السريع.

4.3.4 التغيرات العرضية(الطارئة):

هي تغيرات طارئة تحدث نتيجة حوادث فجائية غالباً لا تكون في الحسبان كالحروب والزلازل والفيضانات والأوبئة .....الخ. ولذا فهي تؤثر على الاتجاه العام للسلسلة الزمنية إما بالزيادة أو النقصان.

ومن الجدير ذكره أن العناصر السابقة الذكر لا تحدث بصورة منفصلة عن بعضها البعض وإنما تحدث جميعاً في وقت واحد، حيث يعمل كل عنصر على التأثير على الظاهرة بدرجة معينة وفي اتجاه معين. بحيث يكون أي تغير يطرأ على الظاهرة هو في الواقع المحصلة لجميع القوى والمؤثرات التي تحيط بالظاهرة. أي أن نموذج السلاسل الزمنية التقليدي يفترض بأن قيم السلسلة Yi هي محصلة العناصر

الأربعة المذكورة: الاتجاه العام T، تأثير التغيرات الموسمية S، تأثير التغيرات الدورية C، تأثير التغيرات الطارئة I وهذه المحصلة تتمثل بقانون الضرب

$$Y_T = T . S . C . I$$

وسوف نستعرض كيفية إيجاد تأثير كل عنصر من العناصر السابقة.

1.1.3.4 دراسة الاتجاه العام

يقصد بدراسة الاتجاه العام هو معرفة اتجاه التغيرات الطويلة الأمد هل هي في تزايد أم في تناقص ومعرفة مقدار التغير السنوي في قيمة حدود السلسلة الزمنية. ويتم تعيين الاتجاه العام في إيجاد خط أو منحنى مناسب يصف حركة السلسلة الزمنية خلال فترة زمنية معينة.

ويمكن تعيين الاتجاه العام بعدة طرق مختلفة نذكر منها:

يمكن تعيين الاتجاه العام بعدة طرق منها:
1- طريقة المربعات الصغرى.
2-طريقة المتوسطات المتحركة.

1.1.1.3.4 طريقة المربعات الصغرى (Least Square Method):

بعد الحصول على المنحنى التاريخي للظاهرة يلاحظ وجود اتجاه عام يأخذ شكل المنحنى، ويمكن الحصول على معادلة هذا المنحنى بطريقة المربعات الصغرى والتي تنص على أن أحسن منحنى هو الذي تكون مجموع مربعات انحرافات القراءات عنه أصغر ما يمكن، وتوجد عدة احتمالات للمنحنى نذكر منها:

1- إذا كان المنحنى على صورة خط مستقيم فتكون معادلته على الصورة:

$$Y = a X + b$$

حيث :

Y : قيم الظاهرة

X : الزمن

b ,a : مقادير ثابتة تتعين من خلال حل المعادلات الطبيعية

$$\sum_{i=1}^{n} Y_i = a\sum_{i=1}^{n} X_i + nb$$

$$\sum_{i=1}^{n} X_i Y_i = a\sum_{i=1}^{n} X_i^2 + b\sum_{i=1}^{n} X$$

وبإيجاد قيمة a , b تتحدد معادلة أحسن خط مستقيم يمثل الظاهرة، ومن هذه المعادلة نحصل على القيم الاتجاهية للظاهرة في السنوات المختلفة.

2- إذا كان المنحني في صورة منحنى من الدرجة الثانية فتكون معادلته على الصورة: $y = ax^2 + bx + c$

حيث a , b , c مقادير ثابتة

ولتبسيط العمليات الحسابية في إيجاد قيم الثوابت a , b , c ننقل الأصل بالنسبة للزمن إلى قيمة من القيم الوسطى بحيث يكون مجموع قيم X الجديدة يساوي صفر.

حيث نقوم بتحديد مركز السلسلة إذا كان عدد القيم فردياً من العلاقة:

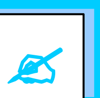
$$t' = \frac{\text{السنة الأولى + السنة الثانية}}{2}$$

بحيث نجعل الزمن يساوي صفراً عند القيمة الوسطى فتكون الفترات التالية لها هي... 1,2,3, والفترات السابقة هي..... -3 , -2 , -1 وبذلك يكون

$$\sum x = 0$$

أما إذا كان عدد القيم زوجياً نأخذ الصفر بالنسبة للزمن بين القيمتين الوسطيتين، فإذا اعتبرنا أن وحدة الزمن تساوي نصف الفترة فإن قيم الزمن في النصف الأعلى من السلسلة هي: ..... -3, -2 , -1 وقيم الزمن في النصف الأسفل من السلسلة هي: .... 1 , 2 , 3 وبذلك يكون $\sum_{i=1}^{n} X_i = 0$

مثال(11.3):



الجدول التالي يعطي إنتاج التبغ في الجمهورية اليمنية

السنة	1994	1995	1996	1997	1998	1999
الإنتاج بالطن(Y)	7170	8069	8574	10229	11469	11081

والمطلوب : إيجاد معادلة الاتجاه العام بطريقة المربعات الصغرى .

الحل :

جدول(9.3)

السنة	X_i	الإنتاج بالطن (Y_i)	X_i^2	$X_i Y_i$
1994	-5	7170	25	-35850
1995	-3	8069	9	-24207
1996	-1	8574	1	-8574
1557	1	10229	1	10229
1998	3	11469	9	34407
1999	5	11081	25	55405
$\sum$	0	56592	70	31410

$$\because \sum_{i=1}^{n} Y = a\sum_{i=1}^{n} X + nb$$

$$56592 = a(0) + 6(b)$$

$$56592 = 6b \Rightarrow b = \frac{56592}{6}$$

$$\Rightarrow b = 9432$$

$$\because \sum_{i=1}^{n} X_i Y_i = a\sum_{i=1}^{n} X_i^2 + b\sum_{i=1}^{n} X$$

$$31410 = a(70) + b(0) \Rightarrow 31410 = 70\ a$$

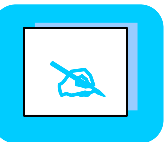
$$\therefore\ a = \frac{31410}{70} = 448.71$$

∴ تكون معادلة الاتجاه العام هي:

$$\hat{Y} = 448.71\ X + 9432$$

مثال (12.3):

الجدول التالي يوضح عدد أجهزة الحاسوب التي باعها أحد المصانع في عدة سنوات .



الربع \ السنة	1999	2000	2001	2002	2003	2004
الأول	34	54	70	90	112	128
الثاني	44	64	90	110	130	150
الثالث	40	60	84	104	120	140
الرابع	30	42	64	80	110	130

المطلوب:

1- احسب معادلة الاتجاه العام للظاهرة الموضحة في الجدول السابق

2- احسب القيم الاتجاه للظاهرة المذكورة .

الحل :

جدول(10.3)

السنة	الربع	X_i	Y_i	X_i^2	$X_i Y_i$
1999	1	-23	34	529	-782
	2	-21	44	441	-924
	3	-19	40	361	-760
	4	-17	30	289	-510
2000	1	-15	54	225	-810
	2	-13	64	169	-832
	3	-11	60	121	-660
	4	-9	42	81	-378
2001	1	-7	70	49	-490
	2	-5	90	25	-450
	3	-3	84	9	-252
	4	-1	64	1	-64
2002	1	1	90	1	90
	2	3	110	9	330
	3	5	104	25	520
	4	7	80	49	560

السنة	الربع	X_i	Y_i	X_i^2	$X_i Y_i$
2003	1	9	112	81	1008
	2	11	130	121	1430
	3	13	120	169	1560
	4	15	110	225	1650
2004	1	17	128	289	2176
	2	19	150	361	2850
	3	21	140	441	2940
	4	23	130	529	2990
	$\sum$	0	2080	4600	11192

$$\because \sum_{i=1}^{n} Y = a\sum_{i=1}^{n} X + nb$$

$$2080 = a(0) + 24\,b \Rightarrow 2080 = 24b$$

$$b = \frac{2080}{24} = 86.7$$

$$\because \sum_{i=1}^{n} X_i Y_i = a\sum_{i=1}^{n} X_i^2 + b\sum_{i=1}^{n} X$$

$$11192 = a(4600) + b(0)$$

$$11192 = 4600\,a + 0$$

$$\Rightarrow a = \frac{11192}{4600} \qquad \Rightarrow a = 2.43$$

∴ معادلة الاتجاه العام هي:

$$\hat{Y}_i = 2.43\ X + 86.7$$

وبالتعويض عن قيم X (‎-1‎، ‎-2‎، ‎-3‎، ....، ‎-23‎، 1، 2، ....، 23) في معادلة الاتجاه العام نحصل على القيم الاتجاهية والموضحة في الجدول (11.3)

جدول(11.3)

السنة	1999	2000	2001	2002	2003	2004
الربع الأول	30.81	50.25	69.69	89.13	108.57	128.01
الربع الثاني	35.67	55.11	74.55	93.99	113.43	132.87
الربع الثالث	40.53	59.97	79.41	98.85	118.29	137.73

الربع الرابع	45.39	64.83	64.27	103.71	123.15	142.59

2.1.3.4 استبعاد أثر الاتجاه العام

بعد إيجاد القيم الاتجاهية اعتماداً على طريقة المربعات الصغرى أو أي طريقة أخرى يتم استبعاد أثر الاتجاه العام وذلك من خلال تقسيم القيم الفعلية (Y_i) على القيم الاتجاهية $\hat{Y}_i$ ، ويكون ناتج القسمة هو الباقي وهو عبارة عن تأثير الموسم والتغيرات الدورية والعشوائية (الطارئة).

مثال(13.3):



اعتماداً على بيانات المثال (11.3) : المطلوب استبعاد أثر الاتجاه العام لبيانات الجدول(9.3).

الحل :

نقوم بحساب القيم الاتجاهية لبيانات المثال المذكور سابقاً حيث معادلة الاتجاه العام هي $\hat{Y}_i = 9432 + 448.71\,X$ ويتم التعويض عن قيمة X (5، 3، 1 ، 1-، 3-، 5-) في معادلة الاتجاه العام ثم نقوم بقسمة قيم Y_i الفعلية على القيم الاتجاه لاستبعاد أثر الاتجاه العام حسب الجدول(12.3)

جدول(12.3)

السنة	X_i	الإنتاج Y_i	$\hat{Y}_i = 9432 + 448.71\,X$	$(Yi / \hat{Y}_i)*100$
1994	-5	7170	7188.45	99.74
1995	-3	8069	8085.87	99.79
1996	-1	8574	8983.29	95.44
1997	1	10229	9880.71	103.52
1998	3	11469	10778.13	106.41
1999	5	11081	11675.55	94.91
$\sum$	0	56592	56592	

وتوضح بيانات الجدول أن الإنتاج متأثر بالتغيرات الدورية والعارضة فقط إذ لا توجد تغيرات موسمية لأن البيانات سنوية.

1.2.3.4 **دراسة التغيرات الموسمية**: لتحديد التغيرات الموسمية للسلسلة الزمنية لابد من معرفة أثر الموسم في السلسلة التي تمثل الظاهرة محل الدراسة ومجموعة

الأعداد التي تبين القيم النسبية للظاهرة خلال شهور السنة تسمى الدليل الموسمي للظاهرة وهناك عدة طرق لحساب الدليل الموسمي نذكر منها :

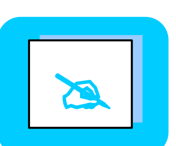
1- طريقة النسبة إلى القيم الاتجاهية : والتي تتلخص في الخطوات الآتية:

أ- الحصول على القيم الاتجاهية للظاهرة

ب- التعبير عن القيم الأصلية لأي ظاهرة في كل فترة زمنية كنسبة مئوية من القيمة الاتجاهية لهذه الفترة وذلك بقسمة القيمة الأصلية (y_i) على القيم الاتجاهية للظاهرة ($\hat{y}_i$) وضرب الناتج في 100

جـ- نحسب متوسطاً مناسباً للنسب المئوية للفترات المتناظرة السابق حسابها في (ب) فنحصل على الدليل الموسمي لهذه الفترة

د- لابد أن يكون مجموع الدليل الموسمي مساويا 1200٪ إذا كانت الفترة بالشهور، 400٪ إذا كانت الفترة ربع سنوية إلخ . وإذا اختلف المجموع عند ذلك فلابد من تعديله بطريقة مناسبة ليصبح المجموع 1200٪، 400٪ حسب الحالة



مثال(14.3):

احسب الدليل الموسمي لبيانات المثال (12.3)

الحل:

نقسم القيم الأصلية للظاهرة في كل فترة Y_i على القيم الاتجاهية لمناظرة ($\hat{Y}_i$) ونضرب الناتج في 100 حيث يتم التعبير عن القيم الأصلية كنسبة مئوية من القيم الاتجاهية لكل فترة فنحصل على الدليل الموسمي كما في الجدول(13.3)

جدول(13.3)

السنة الربع	1999	2000	2001	2002	2003	2004	المتوسط	الدليل الموسمي
الأول	110.35	107.46	100.44	100.98	103.16	99.99	103.73	104
الثاني	123.35	116.13	120.72	117.03	114.61	112.89	117.46	117
الثالث	98.69	100.05	105.78	105.21	101.45	101.65	102.14	102
الرابع	66.09	64.78	75.95	77.14	89.32	91.17	77.41	77
المجموع							400.74	400
المتوسد							100.1	100

ط							9	

واضح من الجدول أن مجموع المتوسطات هو 400.74 لذلك يجب تعديله ليصبح 400٪ للحصول على الدليل الموسمي.

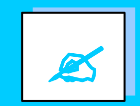
2.2.3.4 استبعاد اثر التغيرات الموسمية:

يتم استبعاد اثر التغيرات الموسمية عن طريق تقسيم القيم الأصلية (الفعلية) للسلسلة الزمنية في كل فترة على الدليل الموسمي المقابل لهذه الفترة فمثلاً القيمة في الربع الأول 1999 هي ناتجة عن تقسيم 34 على الدليل الموسمي للربع الأول 104وذلك اعتماداً على بيانات المثال (12.3) وهكذا .

جدول(14.3)

الربع \ السنة	1999	2000	2001	2002	2003	2004
الأول	32.69	51.92	67.30	86.54	107.69	123.08
الثاني	37.61	54.70	76.92	94.02	111.11	128.21
الثالث	39.22	58.82	82.35	101.96	117.64	134.25
الرابع	38.96	54.55	83.12	103.89	142.85	168.83

مثال(15.3):



كان عدد الركاب الذين تم نقلهم من قبل إحدى الشركات العاملة في النقل الجوي على مدار 5 سنوات والموضحة في الجدول(15.3) وحسب فصول السنة حيث وحدة عدد الركاب هي مائة راكب.

جدول(15.3)

الربع \ السنة	2001	2002	2003	2004	2005
الأول	70	80	85	89	100
الثاني	110	110	120	125	130
الثالث	160	140	160	165	170

جامعة العلوم والتكنولوجيا

الرابع	60	80	90	100	105

المطلوب:

1- حساب القيم الاتجاهية للظاهرة.

2- إيجاد الدليل المسمي للظاهرة.

3- استبعاد أثر التغيرات الموسمية من البيانات.

4- حساب عدد الركاب الذين يتم نقلهم من قبل الشركة في الربع الثالث لسنة 2006.

الحل:

1-نحسب القيم الاتجاهية للظاهرة باستخدام طريقة المربعات الصغرى.

جدول(16.3)

السنة	الربع	X_i	Y_i	X_i^2	$X_i Y_i$
2001	1	-19	70	361	-1330
	2	-17	110	289	-1870
	3	-15	160	225	-2400
	4	-13	60	169	-780
2002	1	-11	80	121	-880
	2	-9	110	81	-990
	3	-7	140	49	980
	4	-5	80	25	-400
2003	1	-3	85	9	-255
	2	-1	120	1	-120
	3	1	160	1	160
	4	3	90	9	270
2004	1	5	89	25	445
	2	7	125	49	875
	3	9	165	81	1485
	4	11	100	121	1100
2005	1	13	100	169	1300
	2	15	130	225	1950
	3	17	170	289	2890
	4	19	105	361	1995
	$\sum$	0	2249	2660	2465

□

$$\because \sum y = a\sum x + nb$$

$$2249 = a(0) + 20\,b \Rightarrow 2249 = 20b$$

$$b = \frac{2249}{20} = 112.45$$

$$\because \sum xy = a\sum x^2 + b\sum x$$

$$2465 = a(2660) + b(0)$$

$$2465 = 2660\,a + 0$$

$$\Rightarrow a = \frac{2435}{2660} \qquad \Rightarrow a = 0.927$$

∴ معادلة الاتجاه العام هي:

$$y = 0.927\,X + 112.45$$

حيث نقطة الأصل منتصف الفترة بين الربع الثاني والثالث لسنة 2003م،

ووحدة الزمن هي: نصف فترة لأن الفترة هي ربع سنوية

وبالتعويض عن قيم X (1- ، 2- ، 3- ، ...، 19- ، 1، 2، ...، 19)

في معادلة الاتجاه العام نحصل على القيم الاتجاهية للظاهرة والموضحة في الجدول التالي:

جدول (17.3)

الربع \ السنة	2001	2002	2003	2004	2005
الربع الأول	95.1	102.4	109.7	117	124.3
الربع الثاني	96.9	104.2	111.5	118.9	126.2
الربع الثالث	98.7	106	113.4	120.7	128.0
الربع الرابع	100.6	107.9	115.2	122.5	129.8

2- لإيجاد الدليل الموسمي للظاهرة : نقوم بتقسيم القيم الأصلية للظاهرة في كل ربع (قيم Y) على القيم الاتجاهية المقابلة ونضرب الناتج في 100، حيث يتم التعبير عن القيم الفعلية للظاهرة كنسبة مئوية من القيم الاتجاهية لكل ربع سنة،

ثم نقسم المتوسط لكل ربع على المتوسط لجميع الأرباع ونضرب في 100 فنحصل على الدليل الموسمي كما هو موضح في الجدول (18.3)

جدول(18.3)

السنة / الربع	2001	2002	2003	2004	2005	المتوسط	الدليل الموسمي
الأول	73.6	78.1	77.5	76.1	80.5	77.16	77.14
الثاني	113.5	105.6	107.6	105.1	103.0	106.96	106.92
الثالث	162.1	132.1	141.9	136.7	132.8	141.12	141.07
الرابع	59.6	74.1	78.1	81.6	80.9	74.86	74.83
$\sum$						400.1	400
المتوسط						100.03	100

3-يتم استبعاد أثر التغيرات الموسمية وذلك بقسمة القيم الفعلية (الأصلية) للظاهرة في كل فترة زمنية على الدليل الموسمي المقابل لهذه الفترة.

والجدول (19.3) يبين قيم الظاهرة بعد تخليصها من أثر التغيرات الموسمية أي بعد استبعاد أثر التغيرات الموسمية منها.

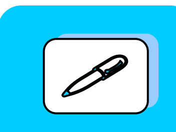
جدول(19.3)

السنة / الربع	2001	2002	2003	2004	2005
الأول	90.70	103.71	110.19	115.37	129.63
الثاني	102.88	102.88	112.23	116.91	212.59
الثالث	113.42	99.42	113.42	116.96	120.51
الرابع	80.18	106.91	120.27	133.64	140.32

4- عدد الركاب الذين يتم نقلهم من قبل الشركة في الربع الثالث من العام 2006م نحصل عليه من التعويض بقيمة $X = 25$ في معادلة الاتجاه العام التي سبق الحصول عليها فنحصل على عدد الركاب الذين سيتم نقلهم من قبل الشركة كما يلي: $\hat{y} = 0.915 \ X + 112.45$

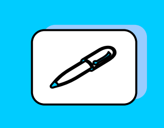
$$\hat{y} = 0.915\ (25) + 112.45 = 135.33$$





1. عرف السلسلة الزمنية ؟
2. أذكر الهدف من استخدام السلسلة الزمنية ؟

تدريب (8)



1- أذكر عناصر السلسلة الزمنية ؟
2- عرف الاتجاه العام واذكر أهم مميزاته.
3- عرف التغيرات الموسمية.



نشاط



عزيزي الدارس : قم بزيارة لوزارة الزراعة واطلب من المختصين هناك تزويدك بمعلومات عن كميات السماد المستخدمة في الأراضي الزراعية وكميات الإنتاج وادرس العلاقة بين كميات الإنتاج الزراعي وكمية السماد المستخدمة..باستخدام الانحدار الخطي .

أسئلة التقويم الذاتي (5) :

أ- الجدول الآتي يعطي كمية استهلاك الماء الصافي في الفترة من (1998-2005) في إحدى المدن :

السنوات	1998	1999	2000	2001	2002	2003	2004	2005
قيم الظاهرة بمئات اللترات (y)	438	454	499	540	585	606	683	813

المطلوب :

1- احسب معادلة الاتجاه العام للظاهرة الموضحة في الجدول بطريقة المربعات الصغرى.

2- احسب الكمية المتوقعة للاستهلاك في سنة 2006م.

ب- كانت مبيعات إحدى الشركات في الفترة من 2003 إلى 2005 وكانت المبيعات ربع سنوية كالآتي:

الربع \ السنة	2003	2004	2005
الأول	20	25	25
الثاني	20	20	15
الثالث	20	25	30
الرابع	35	35	30

المطلوب:

1- حساب القيم الاتجاهية للظاهرة.

2- إيجاد الدليل المسمي للظاهرة.

3- استبعاد أثر التغيرات الموسمية من البيانات.

4- قدر مبيعات الشركة في الربع الثالث لسنة 2006.

5. الخلاصة

تناولت هذه الوحدة دراسة العلاقة بين الظواهر المختلفة سواءً كانت اجتماعية أو اقتصادية... أي من حيث نوع المتغيرات المدروسة كمية أو وصفية حيث يمكن تقسيم هذه المتغيرات إلى متغيرات مستقلة وأخرى تابعة، والمتغيرات التابعة هي التي تتأثر بغيرها فتتغير تبعاً لتغير غيرها، ولهذا سميت متغيرات تابعة.أما المتغيرات المستقلة فهي التي تتغير تبعاً لقوانين وقواعد خاصة بها، ولا علاقة لها بغيرها من المتغيرات.

وإذا أمكن تمثيل العلاقة بين المتغير التابع والمتغير المستقل بخط مستقيم سميت علاقة خطية، أما إذا كان هذا الخط الذي يمثل هذه العلاقة منحنياً سميت علاقة غير خطية.وقد تكون العلاقة طردية إذا كان المتغيران يتغيران بالاتجاه نفسه وقد تكون عكسية عندما يتغيران باتجاهين متعاكسين.وذلك من خلال دراسة معامل الارتباط الخطي البسيط ومعامل ارتباط الرتب وتحديد نوع العلاقة من حيث القوة أو الاتجاه.

ونستعرض فيما يلي بعض الصيغ الرياضية التي تم تناولها في الوحدة:

- معامل الارتباط الخطي البسيط الذي يحدد قوة العلاقة بين متغيرين كميين والذي تتراوح قيمته بين (± 1) من خلال عدة صيغ منها:

$$r=\frac{\sum X_i Y_i - n\bar{X}\bar{Y}}{\sqrt{\left(\sum X_i^2 - n\bar{X}^2\right)\left(\sum Y_i^2 - n\bar{Y}^2\right)}}$$

- معامل ارتباط (سبيرمان) والذي يسمى أيضاً معامل ارتباط الرتب والذي يستخدم عندما يكون المتغيرين مقاسين بمقياس ترتيبي ،كما يستخدم عندما يكون كلا المتغيرين مقاساً بمقياس فئوي بعد أن نستبدل قيم المتغيرين برتب. وأن الاختلاف بين النتيجتين اللتين نحصل عليهما باستخدام المعاملين لا تزيد عن 2٪. ويحسب من العلاقة:

$$r_{rank}=1-\frac{6\sum D^2}{N(N^2-1)}$$

كما استعرضنا في هذه الوحدة معامل الاقتران والذي يستخدم لقياس قوة الارتباط بين متغيرين نوعيين ولكل متغير صفتين فقط ويحسب من:

$$r_c = \frac{n_{11} \cdot n_{22} - n_{21} \cdot n_{12}}{n_{11} \cdot n_{22} + n_{21} n_{12}}$$

وناقشنا أيضاً معامل التوافق الذي يستخدم لإيجاد العلاقة بين الصفات أو الظواهر التي تنقسم أحداهما أو كلتاهما إلى أكثر من صفتين وقيمته دائماً محصورة بين الصفر والواحد الصحيح. ويحسب من الصيغة:

$$r_c = \sqrt{\frac{G-1}{G}}$$

وتناولت الوحدة أيضاً تحليل الانحدار الذي هو واحد من الأدوات الإحصائية الأكثر استخداماً. فهو يعطينا طريقة سهلة وبسيطة لتحديد العلاقة بين المتغيرات بمعادلة رياضية تحتوي على متغير تابع Y مع واحد أو أكثر من المتغيرات المستقلة.

و من أكثر الطرق شيوعاً في تقدير معالم نموذج الانحدار الخطي $\propto, \beta$ هي طريقة المربعات الصغرى، والتي تجعل مجموع مربعات الانحرافات (الأخطاء العشوائية) أصغر ما يمكن. وذلك من الصيغ التالية:

$$\hat{\beta}_i = \frac{n\sum X_i \; y_i - \sum X_i \sum y_i}{n\sum X_i^{\;2} - \left(\sum X_i\right)^2}$$

$$\hat{\beta}_o = \bar{Y} - \hat{\beta}_i \bar{X}$$

ويكتب نموذج الانحدار بصورته النهائية بالصيغة التالية: $\hat{Y} = \hat{\propto} + \hat{\beta} X + e$

كما استعرضت الوحدة السلاسل الزمنية التي تعرف بأنها عبارة عن مجموعة القراءات التي تأخذها ظاهرة ما في فترات زمنية متعاقبة غالباً ما تكون متساوية. تتكون السلسلة الزمنية من العناصر الآتية:

1. الاتجاه العام: يشير إلى الحركة المنتظمة التي تعكس النمو أو الركود أو التناقص على مدى فترة طويلة.
2. التغيرات الموسمية: تشير إلى متوسط التغير المنتظم بصفة دورية خلال سنة أو أقل.
3. التغيرات الدورية :تشير إلى التحركات الدورية التي تتكرر صعوداً أو هبوطاً -أعلى أو أسفل - خط الاتجاه العام.

التغيرات الطارئة: تحدث نتيجة لأسباب غير متوقعة أو غير منتظمة ولا يمكن التحكم فيها كالزلازل والفيضانات.

كما ذكرنا أن الهدف من الدراسة الإحصائية للسلاسل الزمنية هو الكشف عن التغيرات التي تطرأ على الظواهر التي ندرسها خلال فترة زمنية معينة حتى يمكن معرفة أنواع هذه التغيرات وقياس كل نوع منها.

يقصد بدراسة الاتجاه العام معرفة اتجاه التغيرات الطويلة الأمد ، هل هي في تزايد أم في تناقص، ومعرفة مقدار التغير السنوي في قيمة حدود السلسلة الزمنية.



6.لمحة مسبقة عن الوحدة الرابعة:

سبق أن تعلمنا في الوحدة الأولى أن علم الإحصاء ينقسم إلى قسمين أساسيين هما: الإحصاء الوصفي ، والإحصاء الاستدلالي الذي سيمكنك من دراسة خصائص ومعالم المجتمع من خلال بيانات العينة والمقاييس الإحصائية المحسوبة منها. وفي هذه الوحدة سوف تتعرف على واحد من أهم التوزيعات الاحتمالية في الإحصاء وهو التوزيع الطبيعي ، وطريقة رسم منحنى التوزيع وحساب المساحات تحت المنحنى .

أما القسم الآخر فيتناول أحد الموضوعات الهامة في الإحصاء الاستدلالي وهو اختبار الفرضيات من حيث صياغتها ، والخطوات الأساسية لاختبارات الفروض، ودور الفروض الإحصائية في اتخاذ القرار ، ومستوى المعنوية.



7. إجابات التدريبات:

تدريب (1):

الارتباط: هو ظاهرة إحصائية تُطلقْ على العلاقة بين متغيرين.

وقد يكون الارتباط:

موجب: ويشير إلى تزايد المتغيرين المستقل والتابع معاً.

سالب: ويشير إلى تزايد في أحد المتغيرات يقابله تناقص في المتغير الآخر.

معامل الارتباط: .هو مقياس إحصائي يدل على مقدار ونوع العلاقة بين المتغيرات سواءً طردية كانت أو عكسية وتتراوح قيمته بين $-1 \leq r \leq 1$.

ومن الأمثلة على ذلك:

1- دراسة العلاقة بين دخل الفرد ومستواه التعليمي .

2- دراسة العلاقة بين ظاهرة الطلاق وتشرد الأطفال.

3- دراسة العلاقة بين السمات الشخصية لموظف التسويق وكمية المبيعات.

تدريب(2):

الحل : نكون الجدول التالي:

i	x	y	$x\,y$	x^2	y^2
1	6	2	12	36	4
2	4	6	24	16	36
3	8	5	40	64	25
4	5	7	35	25	49
5	7	5	35	49	25
$\sum$	**30**	**25**	**146**	**190**	**139**

نستخدم الصيغة التالية للحصول على معامل الارتباط

$$r = \frac{n\sum X_i Y_i - \sum X_i \sum Y_i}{\sqrt{\left(n\sum X_i^2 - \left(\sum X_i\right)^2\right)\left(n\sum Y_i^2 - \left(\sum Y_i\right)^2\right)}}$$

$$r=\frac{(5)(146)-(30)(25)}{\sqrt{\left(5(190)-(30)^2\right)\left(5(139)-(25)^2\right)}}$$

$$\therefore r=\frac{730-750}{\sqrt{(950-900)(695-625)}}=\frac{-20}{\sqrt{(50)(70)}}$$

$$\therefore r=\frac{-20}{\sqrt{3500}}=\frac{-20}{59.161}=-0.338$$

النتيجة الأخيرة تشير إلى وجود علاقة ضعيفة وعكسية بين الظاهرة X والظاهرة Y أي أنه كلما زادت قيمة X نقصت قيمة Y .

تدريب(3):

أهم خصائص معامل الارتباط

1- يأخذ معامل الارتباط r قيمة من القيم المحصورة بين(1+ , 1-).

2- تكون قيمة معامل الارتباط r (1 ±) عندما تكون العلاقة بين المتغيرين تامة وعندما لا تساوي أي منها تكون غير تامة.

3- في حال عدم وجود علاقة بين المتغيرين فإن قيمة معامل الارتباط تساوي الصفر

4- تكون العلاقة طردية عندما يكون معامل الارتباط موجب وتكون عكسية عندما يكون معامل الارتباط سالب.

وأهم مميزاته: أن قيمة معامل الارتباط مجردة من وحدات القياس .

ومن أهم عيوب معامل الارتباط لبيرسون هي أنه لا يعبر عن متانة العلاقة بشكل صحيح إلا إذا كانت خطية ، أما إذا كانت العلاقة عبارة عن منحنى فقيمته لا تعبر عن متانة العلاقة.

تدريب(4):

نلاحظ أن بيانات الظاهرتين في جزء الجانب العملي والجانب النظري هي رتب لتقديرات الطلاب ولذا لا تحتاج هذه البيانات إلى ترتيب تصاعدي أو تنازلي لتحديد الرتب.

i	رتب درجات المعمل (X)	رتب درجات المحاضرات (Y)	D	D^2
1	8	9	-1	1
2	3	5	-2	4
3	9	10	-1	1
4	2	1	1	1
5	7	8	-1	1
6	10	7	3	9
7	4	3	1	1
8	6	4	2	4
9	1	2	-1	1
10	5	6	-1	1
$\sum$	--	---	--	24

من الجدول نلاحظ أن:

$$\sum D^2 = 24$$

$$r_{rank} = 1 - \frac{6\sum D^2}{N(N^2 - 1)}$$

وبالتعويض في صيغة (سبيرمان) نحصل على:

$$= r_{rank} = 1 - \frac{6(24)}{10(100-1)} = 1 - \frac{144}{990}$$

$$= 1 - 0.1455 = 0.855$$

تشير نتيجة التحليل إلى أن العلاقة بين مستوى أداء الطلاب في جزء المعمل (x) وجزء المحاضرات (y) طردية وقوية.

تدريب(5):

$$G = \frac{(65)^2}{(92)(70)} + \frac{(5)^2}{(8)(70)} + \frac{(27)^2}{(92)(30)} + \frac{(3)^2}{(8)(30)}$$

$$= 1.0023288$$

ومن الصيغة

$$r_c = \sqrt{\frac{G-1}{G}} = \sqrt{\frac{1.0023288-1}{1.0023288}}$$

$$= \sqrt{\frac{0.0023288}{1.0023288}} = 0.048$$

تشير النتيجة إلى وجود علاقة ضعيفة جداً بين الحالة التعليمية للطالب وجنسه .

تدريب(6):

من خلال النتائج المعطاة في التدريب نستطيع حساب معادلة خط انحدار Y على X نتبع الخطوات التالية:

نوجد المتوسط الحسابي لكل من y و X

$$\bar{X} = \frac{\sum x_i}{n} = \frac{81}{9} = 9$$

$$\bar{Y} = \frac{\sum y_i}{n} = \frac{1368}{9} = 152$$

نوجد قيمة المقدر $\hat{\beta}$ من العلاقة التالية:

$$\hat{\beta} = \frac{\sum X_i Y_i - n\bar{X}\bar{Y}}{\sum X_i^2 - n\bar{X}^2}$$

وبالتعويض في المعادلة السابقة نحصل على ميل الانحدار $\hat{\beta}$

$$\hat{\beta}_i = \frac{12668 - 9\,(9)(152)}{(789) - 9\,(9)^2}$$

$$\hat{\beta} = \frac{356}{60} = 5.93$$

وبالتعويض أيضاً بقيمة $\hat{\beta}$ في الصيغة التالية:

$$\hat{\alpha} = \hat{Y} - \hat{\beta}\,\bar{X}$$

$$\hat{\alpha} = 152 - (5.93)*9$$

$$\hat{\alpha} = 98.63$$

وبالتالي فإن معادلة الانحدار التقديرية هي: $\hat{Y} = 98.63 + 5.93\,x$

تدريب(7):

1- تعرف السلسلة الزمنية بأنها عبارة عن مجموعة القراءات التي تأخذها ظاهرة ما في فترات زمنية متعاقبة غالباً تكون متساوية.

2- الهدف من الدراسة الإحصائية للسلاسل الزمنية هو الكشف عن التغيرات التي تطرأ على الظواهر التي ندرسها خلال فترة مدة معينة حتى يمكن معرفة أنواع هذه التغيرات وقياس كل نوع منها.

تدريب(8):

1. تتكون السلسلة الزمنية من العناصر الآتية:

1- الاتجاه العام .

2- التغيرات الموسمية .

3- التغيرات الدورية.

4-التغيرات الطارئة (العشوائية).

2. الاتجاه العام:

هو العنصر الذي يقصد به الحركة المنتظمة للسلسلة الزمنية عبر فترة زمنية طويلة نسبياً ويعتبر في العادة أهم عناصر السلسلة الزمنية وغالباً ما يعتمد كعنصر وحيد في بناء التوقعات.

وأهم ما يميز الاتجاه العام هو أن التغير الذي يطرأ عليه يكون تدريجي وليس مفاجئ ويظهر أثره واضحاً بعد تراكمه فترة طويلة الأمد حيث تستمر هذه التغيرات في نفس الاتجاه لفترة طويلة وإذا ما غيرت اتجاهها فإن الاتجاه الجديد يستمر لفترة طويلة أيضاً.

3. **التغيرات الموسمية:**

هي التغيرات ذات الطبيعة الزمنية الدورية والتي تتكرر بانتظام خلال فترة زمنية لا يزيد طولها على السنة، فقد تكون يومية أو أسبوعية أو شهرية أو فصلية أو سنوية. أي أنها التغيرات المتشابهة التي تظهر في الأسابيع أو الأشهر أو الفصول المتناظرة خلال الفترات الزمنية المختلفة التي تعود إليها مشاهدات السلسلة الزمنية.

8. قائمة المصطلحات:

- **الارتباط** Correlation : هو ظاهرة إحصائية تطلق على العلاقة بين متغيرين.
- **معامل الارتباط الخطي** Linear Coefficient of Correlation: هو مقياس إحصائي يستخدم لقياس قوة واتجاه العلاقة بين متغيرين كميين فقط بينهما علاقة خطية.
- **شكل الانتشار** Scatter Diagram : لوحة تستخدم لتمثيل العلاقة بين متغيرين باستخدام الرسم البياني وتعطي تصور فيما إذا كانت العلاقة بين المتغيرين (X و Y) خطية أم غير خطية وأيضاً قوة هذه العلاقة واتجاهها.
- **معامل ارتباط الرتب(سبيرمان)** Rank Correlation Coefficient : مقياس إحصائي يستخدم لقياس قوة العلاقة بين رتب متغيرين وصفيين أو متغير وصفي وآخر كمي.
- **الانحدار الخطي** Linear Regression: معادلة خط مستقيم تمثل العلاقة بين متغير تابع y ومتغير مستقل x لتقدير ميل المستقيم والمقدار الثابت.
- **معامل التوافق** Coefficient of Contingency: مقياس إحصائي يقيس قوة واتجاه العلاقة بين متغيرين وصفيين كل منهما له تصنيفين أو أكثر.
- **معامل الاقتران** Coefficient of Association : مقياس إحصائي يستخدم لقياس متانة واتجاه العلاقة بين متغيرين لكل متغير تصنيفين فقط.
- **السلاسل الزمنية** Time Series: هي عبارة عن بيانات عن ظاهرة ما في فترات زمنية متعاقبة قد تكون سنوية أو شهرية أو ....الخ. وتستخدم بغرض الكشف عن التغيرات التي قد تطرأ على الظاهرة تحت الدراسة.
- **المربعات الصغرى** Least Square Method : إحدى الطرق الإحصائية الهامة وتستخدم لتقدير معادلة الاتجاه العام للسلسلة الزمنية، كما تستخدم لتقدير معالم نموذج الانحدار.

9.المراجع العربية والأجنبية:

1.9 المراجع العربية

1. أبو صالح، محمد صبحي.(2001): الطرق الإحصائية . الطبعة الثانية دار البازوري العلمية، عمان: الأردن.
2. البلداوي، عبد الحميد .(1997) : الإحصاء للعلوم الإدارية والتطبيقية . الطبعة العربية الأولى، دار الشروق، عمان : الأردن.
3. توفيق، عبد الجبار. (1983): التحليل الإحصائي في البحوث التربوية والنفسية والاجتماعية الطرق اللامعلمية . الطبعة الأولى، مؤسسة الكويت : الكويت .
4. رمضان، زياد .(1997) : مبادئ الإحصاء الوصفي والتطبيقي والحيوي.الطبعة الرابعة، دار وائل للنشر، عمان : الأردن.
5. العواد، منذر حسين.(2003): مبادئ الإحصاء. الطبعة الأولى، مركز الأمين للنشر، صنعاء: اليمن.
6. القاضي، دلال وسهيلة والبياتي، محمود. (2004): الإحصاء للإداريين والاقتصاديين. دار الحامد للنشر والتوزيع، عمان: الأردن.
7. مصطفى، مدني دسوقي.(1968): مبادئ في نظرية الاحتمالات والإحصاء الرياضي وتطبيقاتها في الاستنتاج الإحصائي . دار النهضة العربية، القاهرة : مصر .
8. المنصوب، محمد عبد الكريم .(1998): مفاهيم أساسية في الإحصاء . الطبعة الأولى، منشورات دار الخبرة، صنعاء: اليمن.
9. الهيتي، صلاح الدين حسين.(2004): الأساليب الإحصائية في العلوم الإدارية.الطبعة الأولى، دار وائل للطباعة والنشر، عمان: الأردن.

2.9 المراجع الأجنبية:

1- Draper N . and Smith H ., Applied Regression Analysis , John Wiley and sons Inc ., London 1990.

2- Freund , J . E ., and Walpole , R .E ., (1987) , Mathematical Statistics , 4th end , Prentice Hell .

3- Hole P.; . Introduction .to Mathematical Statistics .New York : John Wiley & Sons, Inc., 1984 .

4- Kendall, M. and Stuart , A.(1977): Advanced Theory of statistics, vol. 1, 4th ed . London : Charles Griffin & Company.

5- Neter, J.; Wassrman , M.H. and Kutner. , Applied Linear Regression Methods . London : Richard D. Irwin, Inc., 1983.

UNIVERSITY OF SCIENCE & TECHNOLOGY
جامعة العلوم والتكنولوجيا

10. التعيينات

السؤال الأول : أختر الإجابة الصحيحة من بين الإجابات الواردة بوضع دائرة حول رقم الإجابة الصحيحة:

1- من أهم أهداف تحليل الانحدار :

1- تحديد العلاقة بين متغيرين

2- التنبؤ بالمتغير التابع Y بدلاله قيم التغير المستقل X

3- أداة للسيطرة والتحكم باتجاه دالة معينة 4- كل ما ذكر

2- إذا كان $\sum_{i=1}^{n} X_i = 210, \sum_{i=1}^{n} Y_i = 174, \sum X^2{}_i = 9100$

$\sum_{i=1}^{n} Y_i^2 = 6280, \sum XY = 7520, n = 6$

فإن معامل الارتباط الخطي البسيط هو:

1) $r = 0.975$

2) $r = 0.935$

3) $r = 0.957$

3- معادلة الانحدار الخطي البسيط تكتب بالصورة :

1) $\hat{Y} = 0.405 + 0.817X_i$

2) $\hat{Y} = 0.85 + 1.817X_i$

3) $\hat{Y} = 0.482 + 0.817X_i$

4-إذا كان $N = 10$, $\sum D^2 = 192.5$ **فإن معامل ارتباط الرتب هو:**

1) $r_{rank} = 1.167$ 2) $r_{rank} = .167$ 3) $r_{rank} = .176$

5-إذا كانت

$\sum X_i = 12$, $\sum Y = 28$, $\sum XY = 108$

$\sum X_i^2 = 47$, $\sum Y_i^2 = 250$, $n = 4$

فإن معامل الارتباط الخطي البسيط بين الظاهرتين X , Y هو:

1) 0.95 2) 0.98 3) 0.96

6- تكون علاقة الارتباط بين متغيرين تامة إذا كانت:

1- $r > 0.90$ 2- $r \geq 0.80$ 3- $r = \pm 1$

7- إذا كانت $N = 10$, $\sum D^2 = 24$ **فإن معامل ارتباط الرتب** r **هو:**

1) 0.585 , 2) 0.855 , 3) .854

8- يمتاز معامل ارتباط سبيرمان للرتب عن معامل ارتباط بيرسون في:

1- إمكانية استخدامه في كل من البيانات الكمية والو صفية

2- سهولة حسابه 3- كل ما ذكر

9- الغرض من إيجاد معادلة خط الانحدار هو:

1- التنبؤ بقيمة المتغير التابع لقيمة محددة من قيم المتغير المستقل

2- قياس متانة العلاقة بين متغيرين

3- بيان نوع العلاقة بين متغيرين

10- إذا كان ميل الانحدار موجباً فإن:

1- الخط المستقيم سيكون اتجاهه إلى الأعلى باتجاه اليمين

2- الخط المستقيم سيكون اتجاهه إلى الأسفل باتجاه اليمين

11- لتقدير معلمات الانحدار في النموذج نستخدم

1- طريقة المربعات الصغرى 2- الانحدار الخطي البسط

3- طريقة المربعات الصغرى الموزونة

12- يستخدم معامل الاقتران إذا كان لدينا متغيرين نوعيين ولكل متغير

1- صفتين فقط 2- صفتين أو أكثر 3- ليس أي ما ذكر

جامعة العلوم والتكنولوجيا UNIVERSITY OF SCIENCE & TECHNOLOGY

السؤال الثاني:

1-أوجد معادلة خط انحدار Y على X والذي يمثل البيانات الآتية:

X	4	5	4	7	8	7	12
Y	4	3	2	9	7	8	11

2- كانت مبيعات إحدى الشركات في الفترة من 2004 إلى 2006 وكانت المبيعات ربع سنوية كالآتي:

السنة / الربع	2004	2005	2006
الأول	13	18	16
الثاني	15	15	17
الثالث	10	14	20
الرابع	24	21	18

المطلوب:

1- حساب القيم الاتجاهية للظاهرة.

2- إيجاد الدليل المسمي للظاهرة.

3- استبعاد أثر التغيرات الموسمية من البيانات.

جامعة العلوم والتكنولوجيا

4

الوحدة الرابعة

مبادئ الإحتمالات واختبارات الفروض



محتويات الوحدة



وقل رب
زدني علماً
جامعة العلوم والتكنولوجيا

1- المقدمــة:

1.1 تمهيـد:

عزيزي الدارس أرحب بك مرة أخرى لدراسة هذه الوحدة والتي تنقسم إلى ثلاثة أقسام رئيسة هي:

1- مقدمة عامة عن الاحتمالات.
2- التوزيع الطبيعي .
3- اختبارات الفروض .

ففي هذه الوحدة سوف نكمل الجزء الثاني من دور الإحصاء والمتعلق باستنتاج معالم المجتمع قيد الدراسة من خلال معطيات نتائج العينة العشوائية المسحوبة منه، ولكي نحقق هذا الهدف فإننا في هذه الوحدة سوف نستعرض مفهوم الاحتمال ودورة في الحياة اليومية من حيث مفهومه الرياضي والإحصائي، وبعض مسلمات الاحتمال وقوانينه أو قواعده الأساسية؛ لما للاحتمالات من استخدامات تطبيقية في جميع نواحي الحياة اليومية، خصوصاً في مجال اتخاذ القرارات الإدارية والاقتصادية.

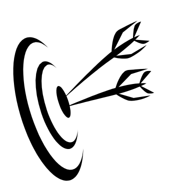
كما سنتناول في هذه الوحدة مفهوم المتغير العشوائي بقسميه (المنفصل والمتصل) عن طريق قوانين الاحتمالات من خلال استعراض أحد أهم التوزيعات الاحتمالية في نظرية الاحتمالات وهو التوزيع الطبيعي(Normal Distribution) من حيث دراسة دالة كثافته الاحتمالية، ورسم منحناه، وأهم خصائصه.وقد حاولتُ في هذا العرض أن أتجنب التعقيدات الرياضية التي قد تكون فوق مستوى الدارس، وكذلك استعرضنا الحالة الخاصة من هذا التوزيع وهي ما يسمى بالتوزيع الطبيعي المعياري، والذي نستطيع من خلاله حساب المساحات تحت المنحنى، كما تجنبنا أيضاً مسألتين بالغتا الأهمية الأولى:هي حساب المساحات باستخدام التكامل في علم الرياضيات، والثانية الجداول اللانهائية للتوزيع الطبيعي عن طريق استخدام جدولاً واحداً بدلاً من هذه الجداول اللانهائية. والذي يستخدم لتقدير معالم المجتمع التي بالا مكان تقديرها بأكثر من طريقة.

ومن خلال مناقشة التوزيع الطبيعي، نكون بذلك قد وضعنا الأساس لتقدير معالم المجتمع التي نستطيع تقديرها بأكثر من طريقة ،ولكن تم الاقتصار في هذا المقرر على اختبار الفرضيات (Hypothesis Testing) الذي يشكل أحد أهم طرق الاستنتاج الإحصائي، ولا يقتصر مجال تطبيقه في الحياة العملية على علم ما بل أن أغلب العلوم العلمية والإنسانية تستخدمه.

2. 1 أهداف الوحدة:

يتوقع منك بعد الانتهاء من دراسة هذه الوحدة أن تكون قادرا على أن:

1- توضح مفهوم الاحتمال
2- تذكر مسلمات الاحتمال
3- تطبق قوانين الاحتمالات على بعض الظواهر.
4- توضح مفهوم المتغير العشوائي
5- تفرق بين المتغير العشوائي المتصل والمتغير العشوائي المنفصل
6- تعرف التوزيع الطبيعي وتحدد دالة كثافته الاحتمالية.
7- ترسم دالة التوزيع الطبيعي بطريقة صحيحة.
8- توضح مفهوم التوزيع الطبيعي القياسي وتحسب المساحة تحت المنحنى الطبيعي من خلال الجدول الخاص بمساحة التوزيع الطبيعي المعياري.
9- توضح مفهوم اختبارات الفروض
10- تذكر الخطوات الأساسية في اختبارات الفروض
11- تبين الفرق بين الخطأ من النوع الأول والخطأ من النوع الثاني.
12- تعرف مستوى المعنوية ثم تحدد ه لكثير من المسائل



3.1 أقسام الوحدة

تم تقسيم هذه الوحدة(مبادئ الاحتمالات واختبارات الفروض) إلى ثلاثة أقسام رئيسة: **تناول القسم الأول:** منها مبادئ الاحتمالات من حيث المفهوم الرياضي للاحتمال وكذلك المفهوم الإحصائي للاحتمال وبعض التعريفات التي تعتمد عليها نظرية الاحتمالات وهذا يرتبط بالهدف الأول.كما تناول هذا القسم مسلمات الاحتمال وكذلك قوانين جمع وضرب الاحتمالات والذي يرتبط بالهدف الثاني من أهداف الوحدة. وتناول أيضاً مفهوم الاحتمال الشرطي ونظرية (بييز) ويرتبط هذا الموضوع بالهدف الثالث.

وفي قسمها الثاني: تناولت هذه الحدة مفهوم المتغير العشوائي والفرق بين المتغيرات العشوائية المتقطعة والمتغيرات العشوائي المتصلة، ومعرفة دالة الكثافة الاحتمالية للمتغير العشوائي المتقطع، ويرتبط هذا الموضوع بالهدفين الرابع

والخامس. أما الهدف السادس فيرتبط بتوضيح مفهوم التوزيع الطبيعي ودالة كثافته الاحتمالية، وخصائص التوزيع الطبيعي، وشكل منحناه، وموقع المتوسط الحسابي للمجتمع في الشكل ويرتبط ذلك بالهدف السابع.

إن التعامل مع هذا التوزيع وحساب المساحات تحت منحنى التوزيع يتطلب معرفة إلى حد معقول بالمفاهيم الرياضية؛ لذلك يتم تحويل قيم المتغيرات الأصلية إلى قيم معيارية حيث يكون لها شكلاً مطابقاً لشكل التوزيع الطبيعي. وهذا ما يطلق عليه التوزيع الطبيعي المعياري، ثم تناولت كيفية حساب المساحات عند ما يأخذ المتغير العشوائي قيمة ما أكبر من أو أصغر من أو محصورة بين قيمتين. بالاعتماد على جداول خاصة بهذا التوزيع وهذا الموضوع يرتبط بالهدف الثامن من أهداف هذه الوحدة.

أما في القسم الثالث والخاص باختبار الفرضيات فقد استعرضنا فيه بعض المفاهيم الإحصائية مثل: الفرضية الصفرية، والفرضية البديلة، والذي يرتبط بالهدف التاسع. أما الهدف العاشر فيرتبط بتوضيح الخطوات الأساسية في اختبار الفرضيات الإحصائية . والخطأ من النوع الأول و الخطأ من النوع الثاني ومستوى المعنوية وهذا يرتبط بالهدفين الحادي عشر والثاني عشر من أهداف الوحدة.

4. 1 القراءات المساعدة:



لكي ترسخ في ذهنك المفاهيم والحقائق التي ستتناولها الوحدة ننصحك بالرجوع إلى القراءات المساعدة الآتية:

1- المنصوب ، محمد عبد الكريم .(1998): مفاهيم أساسية في الإحصاء . الطبعة الأولى ، منشورات دار الخبرة ، صنعاء: اليمن.

2- الهيتي، صلاح الدين حسين.(2004): الأساليب الإحصائية في العلوم الإدارية.الطبعة الأولى، دار وائل للطباعة والنشر، عمان: الأردن.

3- أبو صالح، محمد صبحي ومروة أحمد.(2005) : مبادئ الإحصاء. الطبعة الثانية،منشورات جامعة القدس المفتوحة، عمان: الأردن.

4- سمور ، خالد قاسم وعبيد ، أحمد جمعة .(1995): الاحتمالات . الطبعة الأولى، دار الفكر، عمان:الأردن.

1 .5 الوسائط التعليمية المساندة:



يفضل أن يكون لديك الأدوات الآتية:

- جهاز الكمبيوتر لعرض الـ CDR الخاص بالمقرر. وأوراق رسم بياني وحاسبة بالإضافة إلى أقلام رصاص ومسطرة .

1 .6 ما تحتاج إليه لدراسة الوحدة:

بعد أن تكون قد هيأت المكان المناسب للدراسة ، فإنك ستحتاج إلى توفير بعض المستلزمات التالية: قلم ، ومسطرة ، وآلة حاسبة وورق رسم بياني لرسم دالة التوزيع الطبيعي وبعض الأوراق للكتابة.

وخلال دراستك لمادة هذه الوحدة ستجد في ثنايا الوحدة أسئلة للتقويم الذاتي الهدف منها هو مساعدتك في مراجعة وتلخيص محتويات الوحدة. كما ستجد بعض التدريبات يجب عليك القيام بحلها ، لأنها ترسخ الأفكار المعروضة في هذه الوحدة كما أنها فرصة لاختبار تعلمك ومدى استيعابك وإلمامك بهذه الوحدة.

وعند احتياجك لبعض التوضيحات أثناء دراستك لهذه الوحدة، فلا تتردد في الاتصال بمرشدك الأكاديمي ، ومناقشته فيها ، فسيكون سعيداً بذلك وسيرحب بك كثيراً.

2.مبادئ نظرية الاحتمالات: Probabilities Theory

توطئة:

كثيراً ما نستخدم لفظ (احتمال) أو عبارات توحي بإمكانية تحقق شيء لسنا متأكدين من حدوثه، وتتضح أهمية نظرية الاحتمالات كونها تقوم بدراسة الأحداث التي يكون وقوعها أو عدم وقوعها غير مؤكد الحدوث بالكامل وعلى ذلك فهي تمدنا بالوسائل التي تمكنا من التنبؤ بإمكانية وقوع الحدث.

وتعتبر نظرية الاحتمالات من الفروع الهامة في علم الرياضيات وقد أتسع نطاق تطبيقها حتى أصبح يشمل كافة العلوم الطبيعية والاجتماعية ومنها الإدارية والاقتصادية وبخاصة عمليات اتخاذ القرار في ظل عدم التأكد. وتستخدم تطبيقات نظرية الاحتمال في عدة مجالات منها عمليات الإنتاج ومراقبة جودة الإنتاج وعمليات التسوق، وأنشطة التخطيط للموارد البشرية كما آن شركات التأمين تعتمد أساسا على نظرية الاحتمالات.

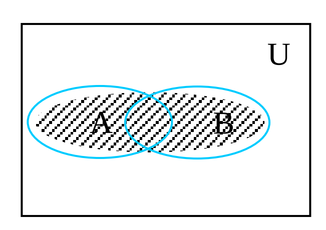
وقبل تعرضنا لمفهوم الاحتمال نقدم بعض التعاريف والمواضيع الأساسية المستخدمة في نظرية الاحتمالات والتي نتعرف من خلالها على القوانين الرئيسة لنظرية الاحتمالات ومنها.

المجموعة(Set): هي أي تجمع من الأشياء أو الأرقام أو الرموز تجمعها خاصية معينة وتكتب بين قوسين أو حاصرتين { } ويدعى أي رمز أو رقم داخل المجموعة عنصراً من عناصر المجموعة مع عدم التكرار. وتكون المجموعة منتهية إذا كان لها بداية ونهاية إذا كانت المجموعة محدودة أي معروف عدد عناصرها مثل مجموع الطلاب جامعة العلوم والتكنولوجيا، وتكون غير منتهية إذا لم يوجد لها بداية ونهاية مثل مجموعة الطيور ومجموعة الأعداد الطبيعية. ويرمز لأي مجموعة بواحدة من بالأحرف الكبيرة ...,A,B,C وعناصرها بالأحرف الصغيرة ...,a,b,c . فعند إلقاء قطعة نقد متزنة فإننا نحصل على(صورة أو كتابة) وتكتب بالصورة:

$$A=\{H\ ,T\}$$

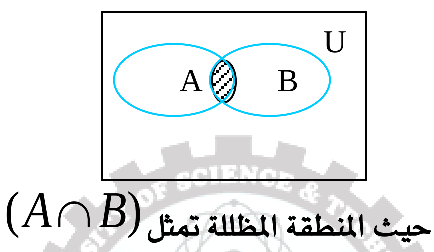
الاتحاد: لتكن $A\ ,B$ مجموعتين فإن اتحاد هما يعطي مجموعة أخرى بحيث أن عناصر مجموعة الاتحاد تنتمي إلى A أو B أو تنتمي إلى كلا المجموعتين ويرمز له بالرمز $A\cup B$ ويمكن تمثله بالشكل الآتي:

UNIVERSITY OF SCIENCE & TECHNOLOGY
جامعة العلوم والتكنولوجيا



حيث المنطقة المظللة تمثل $(A \cup B)$

التقاطع(Intersection): لتكن A , B مجموعتين فإن تقاطعهما عبارة عن مجموعة العناصر المشتركة بين A , B ويرمز له بالرمز $A \cap B$ ويمكن تمثيله بالشكل



حيث المنطقة المظللة تمثل $(A \cap B)$

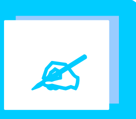
1.2 طرق العد (Numeration Methods):

إن تحديد فضاء العينة يعتبر من الأمور الهامة التي يحتاجها الإحصائي في دراسة الاحتمالات وبالتالي حساب الاحتمال، ولهذا الغرض سوف نتطرق إلى بعض طرق العد التي تساعدنا في الحصول على عدد النتائج الممكنة لتجارب عشوائية.

1.1.2. مبدأ الإضافة:

إذا أجريت تجربة ما بعدد من الطرق مقدارها a وتجربة أخرى بعدد من الطرق b وكانت كلا التجربتين متنافيتي الوقوع فإن كلا التجربتين يمكن إجراؤهما بعدد $a + b$ من الطرق.

مثال(1.4):



إذا رميت قطعة نقد مرة واحدة فكم عدد الطرق التي يمكن أن تظهر بها صورة أو كتابة .

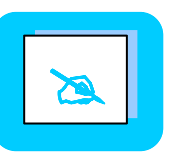
الحل:

هنا توجد حالتان ممكنة الوقوع إما أن تظهر الصورة أو الكتابة وبالتالي فإن 1+1=2 هي عدد الطرق التي يمكن أن تظهر بها صورة أو كتابة.

2.1.2. مبدأ الضرب (Multiplication Rule):

إذا كانت n هي إمكانيات إجراء عملية معينه وبعد الانتهاء من تلك العملية كان هناك m إمكانيات إجراء عملية أخرى فكلا العمليتين يمكن إجراؤهما معاً بعدد $n \cdot m$ من الطرق.

مثال(2.4) :



إذا رمينا قطعة نقد مرتين فحدد عدد عناصر فراغ العينة لهذه التجربة ؟

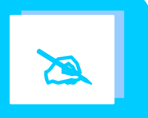
الحل:

في الرمية الأولى تكون النواتج الممكنة هي $a=2$

في الرمية الثانية - - - - - $b=2$

وبالتالي فإن عدد عناصر فضاء العينة هو: $n \cdot m = 2 * 2 = 4$

مثال(3.4):



يمكن لزائر أن يزور مدينة عدن براً أو جواً أو بحراً وبعد أن ينتهي من زيارته لمدينة عدن يمكنه زيارة العاصمة صنعاء براً أو جواً . بكم طريقة مختلفة يمكن لهذا الزائر إتمام زيارته لمدينة عدن والعاصمة صنعاء ؟

الحل:

يمكن أن تتم المرحلة الأولى بثلاث طرق

والمرحلة الثانية يمكن أن تتم بطريقتين

عدد الطرق الكلية المختلفة = 3 × 2 = 6 طرق مختلفة.

ويمكن توضيح ذلك من خلال الشكل (1-5)

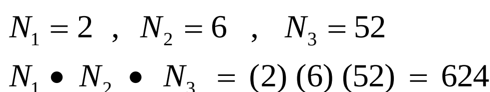
شكل (1-5)

جوا ← بر، جو
برا ← جو، بر
بحرا ← بر، جو

مثال (4.4):

إذا رميت قطعة نقد وحجر نرد ، وسحب كرت من (كوتشينة) ، فحدد عدد عناصر فضاء العينة في هذه التجربة.

الحل:  عدد عناصر العينة هو



$$N_1 = 2 \quad , \quad N_2 = 6 \quad , \quad N_3 = 52$$

$$N_1 \bullet N_2 \bullet N_3 = (2)\,(6)\,(52) = 624$$

3.1.2 التباديل (Permutations):

هي عملية ترتيب أو تنظيم كل أو جزء من مجموعة من الأشياء، وأن تباديل n من الأشياء مأخوذة جميعها هو n! ولتوضيح ذلك نقوم بعرض المثال الآتي:

مثال (5.4) كم عدد الطرق التي يتمكن بها ثلاثة أشخاص من عمل طابور أمام شباك الجوازات؟

الحل:

إذا رمزنا للأشخاص الثلاثة بالرموز C , B, A فإن الموقع الأول يمكن شغره من أي من الأشخاص الثلاثة . فإذا ما تم شغره فإن الموقع الثاني يمكن شغره من أي من الشخصين المتبقيين. وإذا تم شغره فإن الموقع الثالث لا يمكن شغره إلا مرة واحدة وبناءً على ذلك فإن عدد الطرق الممكنة لترتيب الأشخاص الثلاثة في الطابور هي 6= 1 × 2 × 3 أي أن تباديل n من الأشياء مأخوذة جميعها هو n! .

أما إذا لم تأخذ جميعها بل جزء منها في عملية التبديل فإن النتيجة ستختلف كثيراً ولتوضيح ذلك نستعرض المثال الآتي:

مثال (6.4):



كم طريقة مختلفة يمكن لبائع في محل للساعات من عرض ست ساعات في رف لا يتسع إلا لأربع ساعات فقط؟

الحل:

الفراغ الأول يمكن شغرة بأي من الساعات الست فإذا تمت العملية بعرض إحدى الساعات الست فإن الفراغ الثاني يمكن شغرة بأي من الساعات الخمس الأخرى وهكذا تكون عدد الطرق تساوي

$(6)\,(5)\,(4)(3) = 360$ طريقة



تعريف

يدعى وضع أي عدد مثل r ($r \le n$) من الأشياء في ترتيب معين بأنه تبديل n من الأشياء مأخوذة r في كل مرة ويرمز له بالرمز $P(n\,,\,r)$

ويحسب من العلاقة

$$P(n,r)=n(n-1)(n-2)...(n-r+1)=\frac{n!}{(n-r)!}$$

نظرية: عدد تباديل n من العناصر والتي يكون من بينها n_1 عنصراً متماثلاً و n_2 عنصراً متماثلاً و... و n_r عنصراً متماثلاً هو:

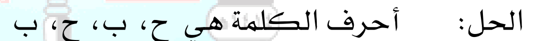
$$\frac{n!}{n_1!.n_2!....n_r!}$$

مثال (7.4):



بكم طريقة مختلفة يمكن ترتيب الحروف الواردة في كلمة (حبحب) مأخوذة مرة واحدة؟

الحل:



أحرف الكلمة هي ح، ب، ح، ب

$$n_1=2 \quad , n_2=2 \;\; , \;\; n=4$$

عدد الطرق $\frac{4!}{2!\,.\,2!}=6$

4.1.2 التوافيق (Combinations) :

التوافيق هي جميع الاختيارات الممكنة بصرف النظر عن الإبدالات

تعريف: توافيق n مأخوذة r في كل مرة هي: $\binom{n}{r}=\frac{n!}{(n-r)!r!}$

مثال(8.4):

بكم طريقة يمكن تشكيل لجنة من ثلاثة أشخاص من مجموعة مكونة من عشرة أشخاص ؟

الحل:

عدد الطرق $\binom{10}{3}=\frac{10\,!}{3!(10-3)!}=120$

فضاء العينية :هو عبارة عن مجموعة المشاهدات التي يمكن ظهورها عند إجراء تجربة ما ويرمز اله بالرمز S.

> يكون الحادثان متساويا الفرصة إذا لم يكن هناك ما يدعو إلى أن نتوقع أن يحدث أحدهما دون الأخر.

الحدث(Event): هو عبارة عن مجموعة جزئية من فضاء العينية . وقد يساوي فضاء العينية نفسه. مثل ظهور الصورة عند رمي قطعة نقود ، أو ظهور الرقم 5 عند رمي حجر النرد مرة واحدة...

الحدث البسيط(Compound Event): إن كل نتيجة من نتائج إجراء تجربة عشوائية تسمى بالحدث البسيط مثل ظهور الصورة عند رمي قطعة نقود مرة واحدة أو ظهور رقم عند رمي حجر نرد.

الحدث المركب(Impossible Event): هو الحادث الذي يتكون من عدة حوادث بسيطة مثل: ظهور عدد فردي عند إلقاء حجر نرد (مركب من الحوادث البسيطة 3, 2 , 1). أو الحصول على مجموع أكبر من ثمانية عند رمي حجري نرد مرة واحدة.

الحدث المستحيل(Sure Event) : هو الحدث الذي يتألف من المجموعة الخالية Φ. مثل ظهور العدد 7 عند رمي حجر نرد مرة واحدة. أو ظهور الصورة والكتابة معاً عند رمي قطعة نقود.

الحدث الأكيد : هو الحدث الذي يساوي فضاء العينية.

الحوادث المتنافية: نقول عن مجموعة من الحوادث بأنها متنافية إذا كان وقوع أحدهما ينفي أو يمنع وقوع الأحداث الأخرى . فعلى سبيل المثال عند رمي قطعة نقد ، فالناتج يكون إما صورة أو كتابة ولا يمكن وقوعهما معاً.والحال كذلك لحادثة ولادة غير توأم فإما أن يكون المولود ذكراً أو أن يكون المولود أنثى.

الحوادث المستقلة: هي الحوادث التي وقوع أحدهما لا يؤثر أو يتأثر بوقوع أو عدم وقوع الحوادث الأخرى. فعند رمي قطعتي نقد مرة واحدة فإما أن يظهر على إحداهما (صورة أو كتابة) لا يمت بصلة لما يمكن أن يظهر على الأخرى.

2.2 الاحتمال (Probabilit) :

للاحتمالات تعريفات كثيرة منها ما هو بسيط معتمد على الإدراك الحسي ومنها ما يعتمد على النظريات المختلفة.

1.2.2 المفهوم الرياضي للاحتمال:

هي الاحتمالات التي يمكن إيجاد قيمها بطرق حسابية دون اللجوء إلى التجارب العشوائية

ويتميز الاحتمال الرياضي عن الاحتمال في معناه العادي في أنه:

1. يمكن قياسه وتحديده بدقة إذا توفرت لدينا معلومات معينة عن الحدث المراد تقدير احتمال وقوعه

2. مقياس موضوعي مستقل عن التقدير الشخصي بمعنى أنه لا يمكن أن يختلف اثنان في قياس الاحتمال الرياضي لوقوع حدث معين إذا كان هذا القياس ممكناً. ويتم تحديد احتمال حدث معين طبقاً لهذا المفهوم بالصيغة:

$$\text{احتمال حدوث الحدث} = \frac{\text{عدد حالات النجاح}}{\text{عدد الحالات الشاملة}}$$

فمثلاً إذا رمينا قطعة نقد مرة واحدة فإن فضاء العينة $S=\{H , T\}$ وإذا كان الحادث A يعني عدد مرات ظهور الصورة فإن الحدث $A=\{H\}$ واحتمال وقوع الحادث $P(A)$ يمكن حسابه من الصيغة:

$$P(A)=\frac{\text{عدد عناصر الحدث A}}{\text{عدد عناصر فضاء العينة S}}$$

$$P(A)=\frac{1}{2}$$

ومن هذا نلاحظ أن احتمال ظهور الصورة يساوي احتمال ظهور الكتابة يساوي $\frac{1}{2}$

مثال(9.4):



احسب احتمال ظهور رقم زوجي عند رمي حجر نرد مرة واحدة.

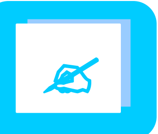
الحل:

النتائج الممكنة التي يمكن الحصول عليها عند رمي حجر النرد مرة واحدة هي:

$$S=\{1 , 2 , 3 , 4 , 5 , 6\}$$

نلاحظ أن فضاء العينة يحتوي على ست عناصر وان الأرقام 6 , 4 , 2 هي أرقام زوجية وهي تمثل عدد حالات النجاح وعليه فإن:

$$P(A)=\frac{3}{6}=\frac{1}{2}$$



مثال(10.4):

صندوق يحتوي على 10 كرات بيضاء و 20 كرة حمراء سحبت كرة من الصندوق. فما احتمال أن تكون الكرة المسحوبة بيضاء ؟ وما احتمال أن تكون حمراء اللون ؟

الحل:

نفرض أن حادث سحب كرة بيضاء هو A ونفرض أن حادث سحب كرة حمراء هو B

$$P(A)=\frac{10}{30}=\frac{1}{3}$$

$$P(B)=\frac{20}{30}=\frac{2}{3}$$

2.2.2 المفهوم الإحصائي للاحتمال:

هو عبارة عن التكرار النسبي لحصول حدث معين عند إجرائه عدداً كبيراً من المرات ونوضح ذلك من خلال المثال الآتي:



مثال(11.4):

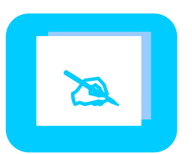
إذا ألقينا قطعة نقد 800 مرة وحصلنا على الوجه كتابة (T) 430 مرة فإن التكرار النسبي لظهور الكتابة يعطي الاحتمال التجريبي لظهور الكتابة هو $P(T)=\frac{430}{800}=0.54$ وإذا ألقينا قطعة النقود 1000مرة أخرى وحصلنا على الوجه كتابة 500 فإن احتمال ظهور الكتابة في الحالتين هو :

$$P(T)=\frac{430+500}{1800}=0.52$$

وهكذا كلما زادت عدد المحاولات فإننا سوف نقترب من الرقم الحقيقي وهو 0.5

UNIVERSITY OF SCIENCE & TECHNOLOGY جامعة العلوم والتكنولوجيا

مثال 12.4:



في تجربة إلقاء حجر نرد مرتين متتاليتين أوجد:

1- فضاء العينة S 2- عدد عناصر S

3- إذا كان الحدث A يعني ظهور عددين مجموعهما 11 على الأقل أوجد الحدث A

4- أوجد P(A)

الحل : 1- فضاء العينة

(6،1)	(5،1)	(4،1)	(3،1)	(2،1)	(1،1)
(6،2)	(5،2)	(4،2)	(3،2)	(2،2)	(1،2)
(6،3)	(5،3)	(4،3)	(3،3)	(2،3)	(1،3)
(6،4)	(5،4)	(4،4)	(3،4)	(2،4)	(1،4)
(6،5)	(5،5)	(4،5)	(3،5)	(2،5)	(1،5)
(6،6)	(5،6)	(4،6)	(3،6)	(2،6)	(1،6)

2- عدد عناصر فضاء العينة = 36 عنصراً.

3- $A = \{ (5,6), (6,5), (6,6) \}.$

4- $P(A) = \frac{3}{36} = \frac{1}{12}.$

3.2.2 خواص (مسلمات) الاحتمال:

بفرض أننا أجرينا تجربة ما ذات فضاء عينة S وكان E أي حادث في فضاء العينة S فإن:

1. مجموع الاحتمالات يساوي الواحد الصحيح أي أن $\sum_{i=1}^{n} P(E_i) = 1$ ، فعلى سبيل المثال في تجربة إلقاء قطعة نقد مرة واحدة إذا رمزنا لحادثة ظهور الصورة بالرمز H ولحادثة ظهور الكتابة بالرمز T فإن احتمال ظهور الصورة مضافاً إليه احتمال ظهور الكتابة يساوي الواحد الصحيح أي أن:

$$P(H) + P(T) = 1$$

ونجمل القول بأن:احتمال وقوع الحادث الأكيد يساوي الواحد الصحيح، وبالرموز :

مسلمات الاحتمال هي: حقائق أو أحكام نسلم بصحتها أو بمشروعيتها من غير حاجة إلى برهان ، وتشكل الأساس الذي يقوم عليه بناء النظرية الاحتمالية كنظرية رياضية.

$$P(S) = 1$$

2. احتمال وقوع الحادث المستحيل يساوي صفراً ، وبالرموز $P(\Phi) = 0$

3.إذا كان الحادث A مجموعة جزئية من فضاء العينية S فإن $0 \le P(A) \le 1$ أي أن الاحتمال دائماً كمية موجبة وتتراوح بين الصفر والواحد الصحيح، أي أنه لا يأخذ كميات سالبة.

4. احتمال حدوث حادث ما $P(A)$ مضافاً إليه احتمال عدم حدوثه $P(A^c)$ يساوي واحد صحيح أي أن $P(A)+P(A^c) = 1$ ويسمى $P(A^c)$ باحتمال الحادث المكمل) المتمم) ويكتب بالصيغة الآتية: $P(A^c) = 1 - P(A)$

تدريب (1)

أعط أمثلة على كل من:



- أحداث مستحيلة وما هو احتمال وقوعها ؟



- أحداث متنافية.
- أحداث مستقلة .



أسئلة التقويم الذاتي (1) :



1) أوجد قيمة كل مما يلي:

$1)\ \dfrac{14!}{15!} \qquad 2)\ \dfrac{(n+3)!}{n!}$

2) كم عدد الطرق التي يمكن أن يجلس بها أربعة أشخاص حول طاولة دائرية؟

3) بكم طريقة يمكن لمدرس أن يختار طالباً أو أكثر من بين خمسة طلاب؟

4) لديك الأرقام 5 ، 4 ، 3 ، 2 ، 1 ، 0 بكم طريقة يمكن تكوين أعداد مكونة من أربعة أرقام ودون تكرار؟

5) اذكر المفهوم الإحصائي للاحتمال.

3.2 بعض القوانين الأساسية في الاحتمالات

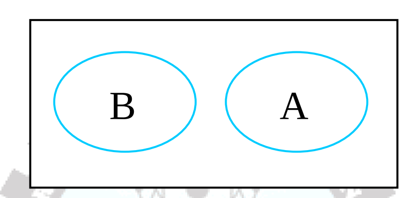
1.3.2 جمع الاحتمالات:

أ. حالة الأحداث المتنافية:

إذا كان A , B حادثان متنافيان أي أن حدوث أحدهما يمنع أو ينفي حدوث الآخر فإن احتمال حدوث أحدهما (A أو B) يساوي حاصل جمع احتماليهما أي أن:

$$P(A \cup B) = P(A) + P(B)$$

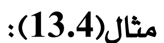
وهذا هو القانون الأول في الاحتمالات والذي يسمى بقانون الجمع.



ويمكن تعميم هذا القانون بالقول أنه إذا كانت $A_1 \ , A_2 , ..., A_n$ حوادث متنافية في فضاء عينة S فإن:

$$P(A_1 \cup A_2 \cup .... \cup A_n) = P(A_1) + P(A_2) + ... + P(A_n)$$





مثال(13.4):

كيس يحتوي على 6 كرات، كرتان سوداء اللون، وكرتان حمراء اللون والباقي صفراء اللون، سحبت كرة من الكيس بطريقة عشوائية. أوجد الاحتمال أن تكون الكرة المسحوبة من اللون الأحمر أو الأصفر.

الحل : بفرض أن الكرة المسحوبة من اللون الأحمر A

وبفرض أن الكرة المسحوبة من اللون الأصفر B

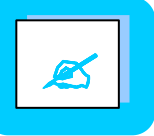
وهما حادثتان متنافيتان وبما أن:

$$P(A \cup B) = P(A) + P(B)$$

$$P(A \cup B) = P(A) + P(B)$$

$$= \frac{2}{6} + \frac{2}{6} = \frac{4}{6} = 0.67$$

مثال(14.4):



بين وجه الخطأ في كل من الجمل الآتية:

1- احتمال أن يفوز سمير في جائزة القرآن الكريم هو 0.85-

2- احتمال أن يفوز أشرف في السباق هو0.89 واحتمال ألا يفوز هو 0.18.

3- احتمال أن يفوز فريق الجامعة لكرة القدم في المباراة القادمة هو 0.76 واحتمال أن يتعادل هو 0.19 واحتمال أن يفوز أو يتعادل هو 0.93 .

الحل:

1- الاحتمال المعطى يتناقض مع مفهوم الاحتمال الذي ينص على أن احتمال أي حادثة غير سالب.

2- الاحتمالات المعطاة تتناقض مع الخاصية الرابعة

$$P(A) + P(A^c) = P(S) = 0.85 + 0.18 > 1$$

3- الاحتمالات تتناقض مع الخاصية (مجموع الاحتمالات يساوي واحد)

$$P(A) = 0.76 \quad , \quad P(B) = 0.19 \quad , P(A \cup \mathbf{B}) = 0.93$$

والحادثتان A , B متنافيتان

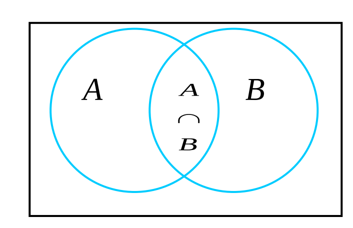
$$P(A \cup B) \neq P(A) + P(B)$$

$$0.93 \neq 0.76 + 0.19$$

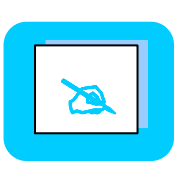
ب. حالة الأحداث غير المتنافية:

إذا كان A , B حادثان غير متنافيين أي أن وقوع أحدهما لا يمنع أو ينفي وقوع الآخر فإن احتمال وقوع أي منهما (A أو B) هو عبارة عن حاصل جمع احتمال كل منهما مطروحاً منه احتمال وقوعهما معاً أي أن:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$



مثال(15.4):



إذا كان A , B حادثين غير متنافيين بحيث أن

$P(A) = 0.4 \ , \ P(B) = 0.7 \ , \ P(A \cap B) = 0.3$ فأوجد احتمال كل

مما يأتي:

1. وقوع أحد الحادثين A أو B
2. عدم وقوع الحادث A

الحل: 1. احتمال وقوع أحد الحادثين A أو B هو:

$$\because \ P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$\therefore \ P(A \cup B) = 0.4 + 0.7 - 0.3 = 1.1 - 0.3 = 0.8$$

2. احتمال عدم وقوع الحادث A = $P(A^c)$ وهو الاحتمال المتمم لوقوع الحدث A وهو عبارة حاصل طرح الاحتمال $P(A)$ من الواحد الصحيح

لأن $\because \ P(A) + P(A^c) = 1$

$$\therefore \ P(A^c) = 1 - P(A)$$

$$\therefore \ P(A^c) = 1 - 0.4 = 0.6$$

تدريب (2)



سحبت ورقة من مجموعة أوراق اللعب . فما هو احتمال أن تكون الورقة المسحوبة تحمل الرقم ثلاثة أو ثمانية أو صورة ؟

تدريب (3)



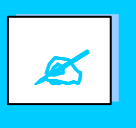
من مجموعة من الطلاب مكونة من 100 طالب رسب 18 طالباً في الإحصاء ورسب 15 طالباً في المحاسبة ورسب 8 طلاب في كلا المادتين. إذا اختير طالباً عشوائياً من هذه المجموعة فما احتمال أن يكون راسباً في الإحصاء والمحاسبة ؟

أ. احتمال الحوادث المستقلة

بفرض أن لدينا الحادثين A, B وكان احتمال وقوع الحادث A مستقلاً عن وقوع الحادث B ولا يتأثر به فإن احتمال وقوع الحادثتين معاً في وقت واحد يساوي حاصل ضرب احتماليهما أي أن:

$$P(A \cap B) = P(A) * P(B)$$

مثال(16.4):



كيس يحتوي على خمس كرات بيضاء وست كرات حمراء. سُحبت كرتان ما احتمال أن تكون الأولى بيضاء والثانية حمراء إذا كان السحب مع الإرجاع.

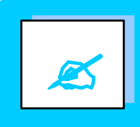
الحل:

احتمال سحب كرة بيضاء مستقل عن احتمال سحب كرة حمراء وشرط الاستقلال هو:

$$P(A \cap B) = P(A) \cdot P(B)$$

$$\therefore P(A \cap B = \left(\frac{5}{11}\right) \cdot \left(\frac{6}{11}\right) = \frac{30}{121} = 0.248$$

مثال(17.4):



كيس يحتوي على 12 كرة ملونة سبع منها حمراء والباقي من اللون الأبيض سُحبت ثلاث كرات من الكيس بطريقة عشوائية واحدة بعد الأخرى مع الإرجاع. أوجد الاحتمال P أن تكون الكرتان الأولى والثانية من اللون الأحمر والكرة الثالثة من اللون الأبيض.

الحل:

الحوادث هنا مستقلة واحتمال وقوع الأولى لا يؤثر في احتمال وقوع الثانية أو الثالثة وبالمثل احتمال وقوع الكرة الثانية لا تؤثر على الأخريات، وعليه فإن:

$$P = \frac{7}{12} \bullet \frac{7}{12} \bullet \frac{5}{12} = \frac{245}{1728} = 0.14$$

مثال(18.4):

إذا كان A , B حادثين في S بحيث أن :

$$P(A)=0.5 \ , P(B)=0.6 \ \ , P(A\cup B)=0.8$$

فأوجد $P(A\cap B)$

الحل:

بما أن الأحداث غير متنافية الوقوع فنستخدم الصيغة:

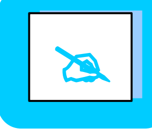
$$P(A\cup B) = P(A)+P(B)-P(A\cap B)$$

$$\therefore \ P(A\cap B)=0.5+0.6-0.8= \ 0.3$$

ب. احتمال الحوادث غير المستقلة

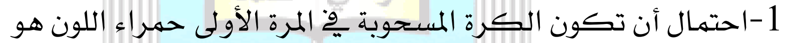
يعتمد ضرب الاحتمالات في هذه الحالة على قاعدة الاحتمال الشرطي، ونستعرض هنا مفهوم الاحتمال الشرطي والذي يعرف بأنه الاحتمال الذي يتم حسابه لحادث ما بعد وقوع حادث آخر قبله ونرمز له بالرمز $p(A/B)$ ويقرأ احتمال الحادث A مع العلم أن الحادث B قد وقع.

مثال (19.4):



بفرض أن لدينا صندوق يحوي 5 كرات حمراء اللون (A) و7 كرات بيضاء اللون (B) فإن:

1-احتمال أن تكون الكرة المسحوبة في المرة الأولى حمراء اللون هو



$$P(A)=\frac{5}{12}$$

2-احتمال أن تكون الكرة المسحوبة في المرة الثانية حمراء اللون دون إعادة الكرة الأولى إلى الكيس مع العلم أن الأولى كانت حمراء هو:

$$P(A/A)=\frac{4}{11}$$

3- احتمال أن تكون الكرة المسحوبة في المرة الثانية حمراء اللون بدون إعادة مع العلم أن الكرة المسحوبة في المرة الأولى كانت بيضاء اللون هو:

$$P(A/B)=\frac{5}{11}$$

4- احتمال أن تكون الكرة المسحوبة في المرة الثانية بيضاء اللون بدون إعادة مع العلم أن الكرة المسحوبة في المرة الأولى كانت حمراء اللون هو:

$$P(B/A)=\frac{7}{11}$$

لاحظ أن :

1. $P(A/B)\neq P(B/A)$ فاحتمال سحب كرة حمراء في المرة الثانية رغم أن الأولى كانت بيضاء هو : $P(A/B)=\frac{5}{11}$ أما احتمال أن تكون الثانية بيضاء مع العلم أن الأولى حمراء هو : $P(B/A)=\frac{7}{11}$

2- إذا كانت الحوادث مستقلة فإن :

$$P(B/A)=P(B), \quad \& \; P(A/B)=P(A)$$

تعريف:

إذا كان لدينا الحادثين A, B غير مستقلين في فضاء العينة S فإن الاحتمال الشرطي للحادثة (A) علماً أن الحادثة (B) قد وقعت يعطي بالعلاقة:

$$P(A/B)=\frac{P(A\cap B)}{P(B)}, \quad P(B)\neq 0 \qquad (❖)$$

وبصورة مماثلة

$$P(B/A)=\frac{P(A\cap B)}{P(A)}, \quad P(A)\neq 0$$

حيث أن:

$$P(A\cap B)=P(B\cap A)$$

وبضرب طرفي المعادلة (❖) ضرباً تناسبياً فإن

$$\left.\begin{aligned} P(A\cap B &= P(B)P(A/B) \\ P(A\cap B) &= P(A)P(B/A) \end{aligned}\right\} \qquad (❖❖)$$

والتي تسمى بالصيغة العامة لضرب الاحتمالات.

مثال (20.4):

في إحدى الكليات نجح 75٪ من الطلاب في امتحان الرياضيات ونجح 85٪ منهم في امتحان البرامج الإحصائية ونجح 90٪ منهم في الرياضيات والبرامج الإحصائية اختير أحد الطلبة بطريقة عشوائية أوجد:

1- إذا كان الطالب راسباً في البرامج الإحصائية فما احتمال أن يكون راسباً في الرياضيات

2- إذا كان الطالب راسباً في الرياضيات فما احتمال أن يكون راسباً في البرامج الإحصائية

3- ما احتمال أن يكون راسباً في الرياضيات أو البرامج الإحصائية

الحل:

نفرض أن حادث رسوب الطالب في الرياضيات $\equiv A$

$$\therefore P(A) = 1 - P(A^c) = 1 - 0.75 = 0.25$$

ونفرض أن حادث رسوب الطالب في البرامج الإحصائية $\equiv B$

$$\therefore P(B) = 1 - P(B^c) = 1 - 0.85 = 0.15$$

ونفرض أيضاً أن حادث رسوب الطالب في الرياضيات والبرامج الإحصائية $\equiv A \cap B$

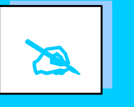
$$P(A \cap B) = 1 - 0.90 = 0.10$$

وبناءً عليه فإن:

$$1)\ P(A/B) = \frac{P(A \cap B)}{P(B)} = \frac{0.10}{0.15} = 0.67$$

$$2)\ P(B/A) = \frac{P(A \cap B)}{P(A)} = \frac{0.10}{0.25} = 0.40$$

$$3)\ P(A \cup B) = P(A) + P(B) - P(A \cap B)$$
$$= 0.25 + 0.15 - 0.10 = 0.30$$



مثال(21.4):

زار رجل عائلة لديها طفلان، فدخل غرفة الضيوف فدخل علية طفل فكان ولداً أوجد الاحتمال P أن يكون الطفل الذي لم يدخل الغرفة ولداً بشرط أن الطفل الذي لم يدخل الغرفة هو الأصغر .

الحل:

نعلم أن فضاء العينية S للعائلات التي لديها طفلان هو :

$$S = \{bb, bg, gb, gg\}$$

حيث: b يعني ولد و g يعني بنت وأيضاً E هي حادثة الأول ولد

$$\therefore E = \{bb\ ,\ bg\}$$

الطفل الذي لم يدخل الغرفة هو الأصغر $A = \{\ bb\ \}$

وحيث أن: احتمال أن يكون الطفل ذكراً هو: $P(b) = \frac{1}{2}$

واحتمال أن يكون الطفل أنثى هو : $P(g) = \frac{1}{2}$

وبناءً عليه، فإن:

$$P(E)=P\left(\{\frac{1}{2}\cdot\frac{1}{2}+\frac{1}{2}\cdot\frac{1}{2}\}\right)=\frac{1}{2}$$

$$P(A)=P\left(\{\frac{1}{2}\cdot\frac{1}{2}\}\right)=\frac{1}{4}$$

وحيث أن:

$$A\cap E=\{bb\}$$

$$\therefore P(A\cap E)=P\left(\{\frac{1}{2}\cdot\frac{1}{2}\right)=\frac{1}{4}$$

احتمال أن يكون الطفل الذي لم يدخل الغرفة ولدا علماً بأنه الأصغر هو:

$$\therefore P(A/E)=\frac{P(A\cap E)}{P(E)}$$

$$=\frac{1/4}{1/2}=\frac{1}{2}=0.5$$

مثال(22.4):

صندوق يحتوي على 12 ليرة ذهب 4 منها عيار 22، الباقي من عيار 21 اختيرت من الصندوق 3 ليرات ذهب واحدة تلو الأخرى دون إرجاع أوجد لاحتمال P أن تكون الثلاث ليرات من عيار 21.

الحل :

$G22$	$G21$
4	8
12	

احتمال أن تكون الليرة المختارة من عيار 21 هو $\frac{8}{12}$ أذن بقي في الصندوق 7 ليرات ذهبية من عيار 21 و 4 ليرات من عيار 22، احتمال أن تكون الليرة الثانية من عيار 21 يساوي $\frac{7}{11}$ إذن بقي في الصندوق 6 ليرات ذهبية من عيار 21و 4

ليرات من عيار 22، واحتمال أن تكون الليرة الثالثة المختارة من عيار 21 يساوي $\frac{6}{10}$ و باستخدام نظرية حاصل الضرب :

$$p = \frac{8}{12} \bullet \frac{7}{11} \bullet \frac{6}{10}$$

$$p = \frac{14}{55}$$



تدريب (4)

ما هو وجه الخطأ في كل مما يلي :

1- $P(A) = 0.48 \quad , \quad P(A^c) = 0.42$

2- $P(B) = 1.02$

3- $P(C) = -0.03$



4.2 نظرية (بييز) (Baye's Theorem):

إذا كانت مجموعة الحوادث $A_1, A_2, ..., A_n$ تمثل تجزيئاً لفضاء العينة S

حيث $i \neq j \;,\; A_i \cap A_j = \phi \;,\; P(A_i) > 0 \;،\; S = \bigcup_{i=1}^{n} A_i$

وهي أحداث منفصلة تعني تقاطعها مثنى مثنى يساوي ϕ وشاملة تعني اتحادها جميعاً يساوي فضاء العينية S وإذا كان B أي حادث في فضاء العينة S بحيث $(P(B) > 0)$ فإن B هي حادثة إتحاد حوادث متنافية أي أن

$$P(B) = \sum_{i=1}^{n} P(A_i \cap B)$$

$$P(B) = \sum_{i=1}^{n} P(A_i)\, P(B/A_i) \qquad (*)$$

وحيث أن

$$P(A_i / B) = \frac{P(A \cap B)}{P(B)}$$

$$\therefore (A \cap B) = P(B)\, P(A_i / B) \qquad (**)$$

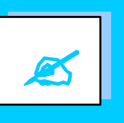
تعتمد نظرية بييز على حساب درجة الاحتمال التقديري الذي يحسب بعد إجراء التجربة ومشاهدة نتائجها.

وسبق أن عرفنا أن $P(A_i / B) = \dfrac{P(A \cap B)}{P(B)}$

وباستعمال قاعدة الضرب والتعويضمن المعادلتين (* ' **) فإن نظرية (بييز) تأخذ الصورة الآتية:

$$P(A_i / B) = \frac{P(A_i)P(B / A_i)}{\sum_{i=1}^{n} P(A_i)P(B / A_i)}$$

مثال(23.4):



ثلاث صناديق متشابهة يحوي الأول 10 كرات 7 منها بيضاء والباقي من اللون الأسود ويحوي الثاني على 5 كرات، 4 منها بيضاء والباقي من اللون الأسود ويحوي الثالث على 5 كرات اثنتان منها بيضاء والباقي من اللون الأسود اختير صندوق من الصناديق الثلاثة بشكل عشوائي أوجد.

1- احتمال سحب كرة بيضاء

2- إذا كانت الكرة المسحوبة بيضاء فما هو احتمال أن تكون من الصندوق الثالث

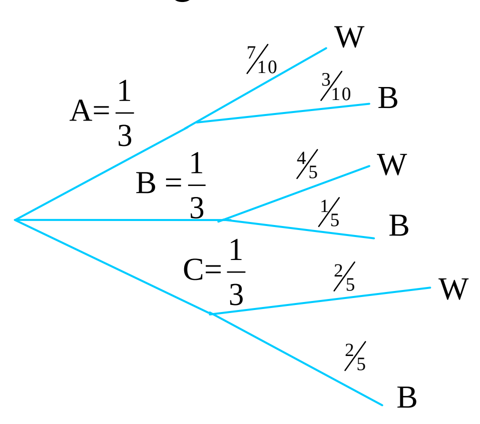
الحل :

سوف نستخدم شجرة الاحتمال كإحدى الطرق المناسبة لحساب احتمال حادث معين إذا كان هذا الحادث أتي من قبل حوادث متعددة تمثل تقسيماً لفضاء العينة، حيث نمثل فضاء العينة بأصل الشجرة ونمثل تقسيم فضاء العينة بفروع الشجرة. لأن شجرة القرار تفيدنا في تمثيل نتائج التجربة الإحصائية المتعددة المراحل، وتعين احتمال كل فرع من فروع الشجرة التي مثلت تلك التجربة.

والآن نفرض أن الصندوق الأول (A) والصندوق الثاني (B) الصندوق الثالث (C)

ونفرض حادث سحب كرة بيضاء (x) و احتمال سحب أي صندوق هو $\frac{1}{3}$

شجرة الاحتمال لتوضيح نظرية بييز



1) $$P(X) = P(A)P(X/A) + P(B)P(X/B) + P(C)P(X/C)$$

$$P(X) = \frac{1}{3} \bullet \frac{7}{10} + \frac{1}{3} \bullet \frac{4}{5} + \frac{1}{3} \bullet \frac{2}{5}$$

$$P(X) = \frac{1}{3}\left[\frac{7}{10} + \frac{4}{5} + \frac{2}{5}\right]$$

$$P(X) = \frac{1}{3}\left[\frac{7+8+4}{10}\right] = \frac{1}{3} \bullet \frac{19}{10}$$

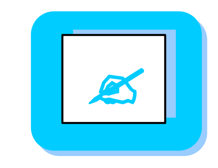
$$P(X) = \frac{19}{30}$$

2) $$P(C/X) = \frac{P(C)P(X/C)}{P(A)P(X/A) + P(B)P(X/B) + P(C)P(X/C)}$$

$$P(C/X) = \frac{\frac{1}{3} \bullet \frac{2}{5}}{\frac{19}{30}} = \frac{2}{15} \bullet \frac{30}{19}$$

$$P(C/X) = \frac{4}{19}$$

مثال (24.4) :



معمل للأحذية يحتوي على ثلاث ماكينات A , B , C تنتج الماكينة A (55٪) من الإنتاج الكلي للمعمل، وتنتج الماكينة B (20٪) من الإنتاج الكلي للمعمل. وتنتج الماكينة C (25٪) من الإنتاج الكلي للمعمل، إذا علم أن نسبة الإنتاج المعيب من الأحذية لهذه الماكينات هي: 5٪، 3٪، 2٪ على التوالي. اختير حذاء من الأحذية بطريقة عشوائية فما احتمال.

1-أن يكون الحذاء معيباً؟

2- أن يكون من إنتاج الماكينة B مع العلم أنه معيباً؟

الحل:

نفرض أن أنتاج وحدة معيبة هو: X وحيث أن:

$$P(A) = 0.55$$
$$P(B) = 0.20$$
$$P(C) = 0.25$$
$$P(X / A) = 0.05$$
$$P(X / B) = 0.03$$
$$P(X / C) = 0.02$$

1)
$$\begin{aligned} P(X) &= P(A)P(X / A) + P(B)P(X / B) + P(C)P(X / C) \\ &= (0.55)(0.05) + (0.20)(0.03) + (0.25)(0.02) \\ &= 0.0385 \end{aligned}$$

2)
$$P(B / X) = \frac{P(B)P(X / B)}{P(A)P(X / A) + P(B)P(X / B) + P(C)P(X / C)}$$
$$P(B / X) = \frac{0.2\ (0.03)}{0.0385}$$
$$P(B / X) = 0.156$$

تدريب (5)



صندوقان يحتويان على عناصر تالفة وأخرى غير تالفة كالآتي:

	الصندوق 1	الصندوق 2
عدد التالف	17	20
عدد غير التالف	3	5



سحب صندوق بطريقة عشوائية وسحب منه عنصر أوجد :

1. احتمال أن العنصر الذي سحب تالف
2. إذا علم أن العنصر تالف فما احتمال أن يكون قد سحب من الصندوق الثاني.

أسئلة التقويم الذاتي (2) :

1- عرف كلاً من الاحتمالات الحسابية ، فضاء العينة ، التجربة العشوائية

2- ما الفرق بين الحوادث المتنافية والحوادث المستقلة مع التوضيح بأمثلة على كل منها.

3- إذا كانت D حادثة أن كتاباً جديداً في الإحصاء سيطبع طباعة ممتازة و E حادثة أنه سيلقى رواجاً في السوق وF حادثة أنه سيتم تبنيه كمقرر جامعي اكتب كلاً من الاحتمالات التالية بصورة رمزية :

ا- احتمال أن الكتاب سيلقى رواجاً ويجري تبنيه كمقرر جامعي

ب- احتمال أن الكتاب لا يطبع طباعة ممتازة ولا يتم تبنيه كمقرر جامعي

ج- احتمال أن الكتاب لا يطبع طباعة ممتازة ويتم تبنيه كمقرر جامعي

د- احتمال أن الكتاب سيطبع طباعة ممتازة ويتم تبنيه كمقرر جامعي

4- في مجتمع معين إذا كان احتمال أن يكون المولود ذكراً يساوي 0.52 وعلى فرض أنه تم تسجيل قيد ثلاث حالات ولادة من المواليد أوجد الاحتمالات الآتية :

أ- أن تكون الحالات الثلاث كلها من الذكور

ب- أن تكون الحالات الثلاث كلها من الإناث.

ج- أن يكون مولودان من الذكور ومولود من الإناث .

د- أن تكون مولودتان من الإناث ومولود من الذكور.

هـ- أوجد مجموع الاحتمالات الحسابية.

و- أن يكون مولود واحد على الأقل من الذكور.

5-اختيرت ثلاثة مصابيح كهربائية بطريقة عشوائية من بين خمسة عشر مصباحاً كهربائياً ، خمسة منها غير صالحة أوجد الاحتمال.

1) جميعها صالحة. 2) واحد فقط غير صالح. 3) واحد على الأقل غير صالح.

3. المتغيرات العشوائية Random Variables :

يعرف المتغير العشوائي بأنه: دالة عددية معرفة على فضاء عينة.
كما يعرف بأنه متغير يأخذ قيماً عددية تحددها نتائج تجربة عشوائية.

يهتم علم الإحصاء بدراسة الظواهر الإحصائية حيث أنها تتغير ولا يمكن معرفة ما سيحدث لها في المستقبل بالضبط وذلك بغرض التنبؤ بهذه الظواهر في المستقبل. وقد لا حظنا في الوحدات السابقة عند تناولنا للتوزيعات التكرارية أن بعض المتغيرات الإحصائية تتصف بكونها غير متصلة (متقطعة) مثل حالة رمي قطعة نقد أو رمي حجر نرد ، والبعض الأخر متصلة (مستمرة) مثل حالة قيم الأطوال والأوزان. وأن هذه التوزيعات التكرارية غالباً ما تكون آتية من عينات مسحوبة من مجتمعات إحصائية معينة. وعندما نكون على معرفة تامة بأن العينة ممثلة للمجتمع الذي سحبت منه نتمكن من بناء كافة أنواع التنبؤات والاستنتاجات بالاعتماد على التوزيع التكراري للعينة مما يستوجب معرفة توزيع هذه المتغيرات التي تحقق فرضيات محددة مرتبطة بطبيعة المجتمع المسحوبة منه العينة ، وهذه الفرضيات تقودنا إلى ما يعرف بالتوزيعات الاحتمالية.

وعندما يكون المتغير العشوائي X عدد محدود من القيمة الحقيقية تحددها نتائج تجربة عشوائية باحتمالات معينة نقول أنه متغير عشوائي مثل رمي قطعة نقد ، وينقسم المتغير العشوائي إلى:

أ. متغير عشوائي منفصل (Discrete Random Variable) : وهو المتغير الذي يأخذ قيماً حقيقية صحيحة في مجال تغيره مثل عدد الطلاب الناجحين في مقرر ما من مجموعة طلاب عددها 20 طالباً. في حين يعرف المتغير العشوائي المتصل: بأنه المتغير الذي يمكن أن يأخذ عدداً غير منته من القيم في مجال تغيره مثل الزمن الذي ينقضي قبل بدأ عمل معين.

1.3 قانون التوزيع الاحتمالي :

يعرف قانون التوزيع الاحتمالي بأنه العلاقة التي تربط بين القيم الممكنة للمتغير العشوائي x واحتمالاتها $P(x)$. أي أن $f(x)$ للمتغير العشوائي X يساوي أو يأخذ القيمة x بحيث أن:

$$f(x) = P(X = x_1) \qquad , \quad x \in S$$

وتسمى هذه الدالة بقانون التوزيع الاحتمالي للمتغير العشوائي X إذا تحقق الشرطان:

$$(1)\quad 0 \le P(x_i) \le 1$$

وهذا الشرط معناه أن الاحتمالات جميعها أكبر من أو تساوي صفر(غير سالب)، أو أقل من أو تساوي الواحد الصحيح.

$$(2)\quad \sum_{i=1}^{n} P(X_i) = 1$$

وهذا الشرط معناه أن مجموع الاحتمالات يساوي الواحد الصحيح. ويطلق على $f(x)$ اسم دالة الاحتمال وهي نفس الصيغة $P(x)$ لهما نفس الدلالة الاحتمالية.

وتوصف هذه العلاقة بالنسبة للمتغيرات العشوائية المنفصلة بالجدول:

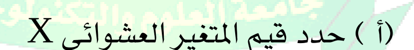
x_i	x_1	x_2	----	x_n
$f(x_1) = p(x_1)$	$p(x_1)$	$p(x_2)$	----	$p(x_n)$

حيث أن

$$P(x) = \frac{\text{عدد عناصر الحدث}}{\text{عدد عناصر فضاء العينة}}$$

مثال (25.4):

في تجربة إلقاء قطعة نقد مرتين إذا كان المتغير العشوائي X يعني عدد مرات ظهور الصورة، أجب عما يلي :


(أ) حدد قيم المتغير العشوائي X

(ب) كون جدول التوزيع الاحتمالي .


الحل :

لاحظ أن الفضاء العيني S لتجربة إلقاء قطعتي نقد معاً هو:

$$S = \{HH, HT, TH, TT\}$$

حيث H صورة، T كتابة، وما يهمنا هنا هو عدد مرات ظهور الصورة فإن :

$$HH \to 2$$
$$HT \to 1$$
$$TH \to 1$$
$$TT \to 0$$

يلاحظ أن المتغير العشوائي x يمكن أن يأخذ أي قيمة من القيم $0 , 1 , 2$

فتكون قيم المتغير العشوائي هي: $X = (0 , 1 , 2)$

جدول التوزيع الاحتمالي يتألف من سطرين أفقيين . السطر الأول يحتوي على قيم المتغير العشوائي X والسطر الثاني يحتوي على الاحتمالات المناظرة لقيم المتغير العشوائي X، بالإضافة إلى عناصر فضاء العينة.

قيم المتغير العشوائي X_i	0	1	2	المجموع
فضاء العينة S	TT	HT, TH	HH	4
الاحتمال $P(x)$	$\frac{1}{4}$	$\frac{2}{4}$	$\frac{1}{4}$	$\frac{4}{4}=1$

لاحظ آن الجدول السابق يشكل جدول توزيع احتمالي حيث أنه يحقق الشرطين:

$$1. \quad 0 \le P(x_I) \le 1$$

$$2. \quad \sum_{I=1}^{n} P(x_I) = 1$$

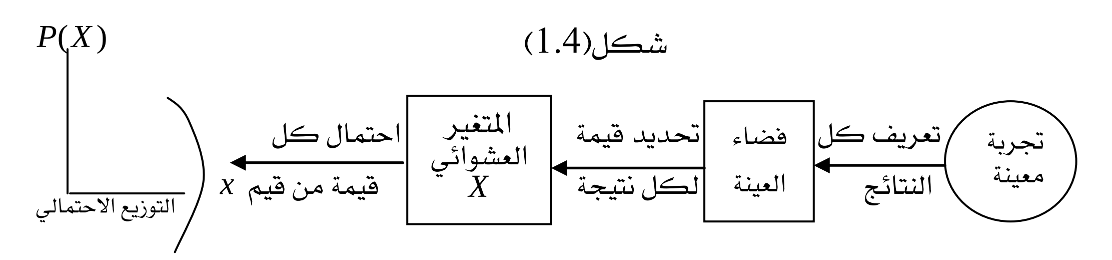
فنلاحظ أن جميع الاحتمالات غير سالبة أي:

$$P(0) = \frac{1}{4} \quad , P(1) = \frac{2}{4} \quad , \quad P(2) = \frac{1}{4}$$

وأن مجموع كل الاحتمالات في التوزيع تساوي واحد صحيح. والشكل التالي(1.4) يعطي تلخيصاً لهذا النموذج، والذي يجعل من الممكن تحديد التوزيع الاحتمالي.

UNIVERSITY OF SCIENCE & TECHNOLOGY

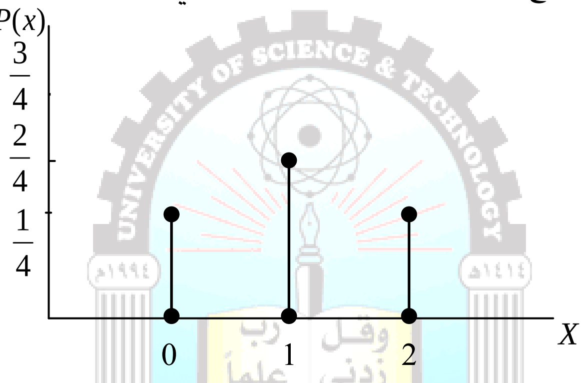
شكل(1.4)



والشكل (2.4) يوضح دالة الاحتمال للمتغير العشوائي الذي يمثل حادثة رمي قطعة نقد مرتين متتاليتين.

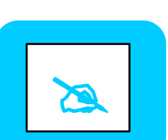
شكل (2.4)

يوضح دالة الاحتمال للمتغير العشوائي للمثال(25.4)



حيث يمثل المحور الأفقي قيم المتغير العشوائي X فيما يمثل المحور الرأسي الاحتمال $p(x)$ والذي يعبر عن التكرارات النسبية.

مثال(26.4):



في تجربة إلقاء حجري نرد إذا كان المتغير العشوائي X يعني الفرق المطلق بين العددين الظاهرين أجب عما يلي :

1- إيجاد القيم التي يأخذها المتغير العشوائي X .

2- كون جدولاً للتوزيع الاحتمالي .

الحل:

1- يمكن إيجاد قيم المتغير العشوائي X من خلال استعراض فضاء العينة والفرق بين عنصري المشاهدة كما يلي:

حجر النرد الأولى

حجر النرد الثانية	1	2	3	4	5	6
1	0 • (1,1)	1 • (1,2)	2 • (1,3)	3 • (1,4)	4 • (1,5)	5 • (1,6)
2	1 • (2,1)	0 • (2,2)	1 • (2,3)	2 • (2,4)	3 • (2,5)	4 • (2,6)
3	2 • (3,1)	1 • (3,2)	0 • (3,3)	1 • (3,4)	2 • (3,5)	3 • (3,6)
4	3 • (4,1)	2 • (4,2)	1 • (4,3)	0 • (4,4)	1 • (4,5)	2 • (4,6)
5	4 • (5,1)	3 • (5,2)	2 • (5,3)	1 • (5,4)	0 • (5,5)	1 • (5,6)
6	5 • (6,1)	4 • (6,2)	3 • (6,3)	2 • (6,4)	1 • (6,5)	0 • (6,6)

UNIVERSITY OF SCIENCE & TECHNOLOGY
١٩٩٤م ١٤١٤هـ
وقل رب زدني علماً
جامعة العلوم والتكنولوجيا

وتوضح الأعداد المرتبطة بالنقاط قيم المتغير العشوائي X للتجربة، والتي تمثل الفرق المطلق بين العددين الظاهرين على وجهي حجري النرد، والتي تراوحت بين أصغر فرق هو (الصفر) وأكبر فرق وهو (5). فتكون قيم المتغير العشوائي هي:

$$X = \{0,1,2,3,4,5\}$$

2- وحيث أن عدد عناصر فضاء العينة بلغ (36) عنصر، فإن جدول التوزيع الاحتمالي يأخذ الصورة الآتية:

X_i	0	1	2	3	4	5
$f(x)$	$\frac{6}{36}$	$\frac{10}{36}$	$\frac{8}{36}$	$\frac{6}{36}$	$\frac{4}{36}$	$\frac{2}{36}$

ب. متغير عشوائي متصل: (Continuous Random Variable):

ذكرنا في البند السابق أن المتغيرات العشوائية المنفصلة تلائم الظواهر التي يمكن تمثيلها بعدد كأعداد العاملين في الكلية، أو أعداد الأطفال في إحدى الأسر. إلا أنه في حالات كثيرة نتعرض لبعض الظواهر المقيَّسة وهي الظواهر التي تمثل الأوزان أو القوى أو الأطوال أو معدل هطول الأمطار أو درجة حرارة جسم ودرجة امتحان الطالب، وقياس مثل هذه المتغيرات هي نقاط على خط اتخذنا عليه تدريجاً أو سلماً للقياس، أي أنها نقاط علي محور الأعداد الحقيقية أو على فترات من هذا المحور، والتي يعجز المتغير العشوائي المنفصل في التعبير عنها، ولذا نلجأ إلى المتغير العشوائي المتصل والذي يعرف على أحداث شاملة تمثل بفترة متصلة، قد تكون محدودة أو غير محدودة .

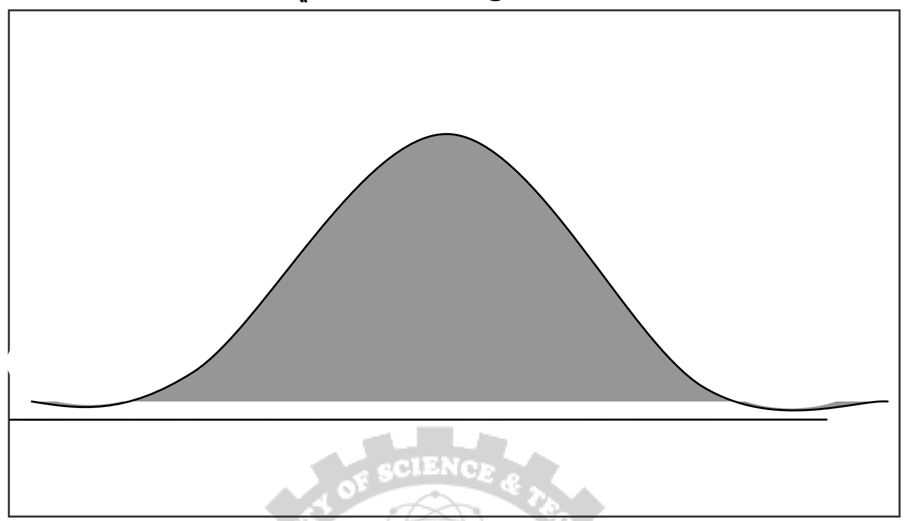
وإذا كنا قد تحدثنا عن التوزيعات الاحتمالية المنفصلة والتي يأخذ المتغير العشوائي فيها قيماً صحيحة منفصلة، فإننا سوف نتناول فيما يأتي بعضاً من التوزيعات الاحتمالية المتصلة التي يكون المتغير العشوائي فيها متصلاً.

2.3 التوزيع الطبيعي (جاوس)(Gauss) (The Normal Distribution) :

يعتبر التوزيع الطبيعي حجر الأساس للنظرية الإحصائية الحديثة وهو أكثر التوزيعات الإحصائية أهمية واستخداماً ويطلق عليه أحياناً " توزيع جاوس " نسبة إلى مكتشفه " كارل . جاوس " الذي نشره عام 1733م . ويسمى أحياناً بالتوزيع الجرسي لأنه يشبه الجرس المعكوس كما أن كثيراً من التوزيعات المهمة تُشتق منه أو تؤول إليه في حالات معينة. كما يعتبر التوزيع الطبيعي متغيراً عشوائياً متصلاً لأن فئته الشاملة تتكون من عدد لا نهائي من القيم الحقيقية التي تختلف عن بعضها البعض بمقادير متناهية في الصغر بحيث يمكن ترتيبها على مقياس متصل. ويمتاز التوزيع الطبيعي بأنه متماثل حول الوسط الحسابي الذي يمثله. ويأخذ المنحنى شكل الناقوس انظر الشكل(3.4)

شكل (3.4)

يوضح المنحنى الطبيعي



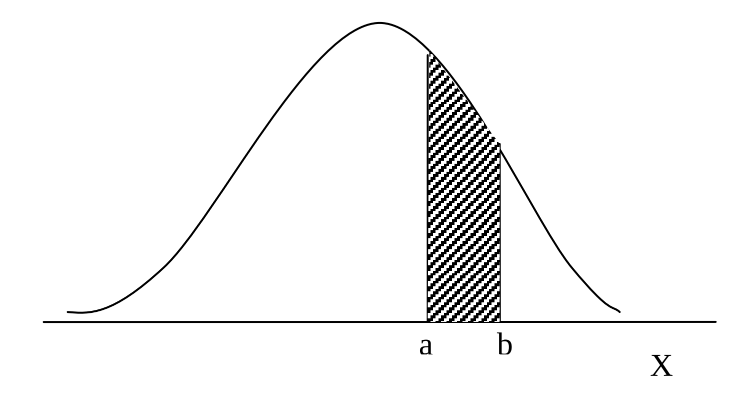
وفي الحياة العملية توجد العديد من المتغيرات العشوائية المتصلة مثل الأطوال والأوزان، والحجوم ،ودرجات اختبار الذكاء وأعمار المصابيح الكهربائية وغيرها من الأمثلة التي يأخذ توزيعها في مجتمع كبير شكل التوزيع الطبيعي(المعتدل أو السوي).

ولقد تبين أن منحنى التوزيع الاحتمالي لأي من المتغيرات العشوائية التي سبق ذكرها وغيرها يقترب شكله من منحنى متصل يشبه الجرس ويسمى بالمنحنى ألاعتدالي " Normal Curve"" . وحيث أن المنحنى الإعتدالي هو منحنى توزيع احتمالي متصل، فإن أية نقطة من النقاط الموجودة على هذا المنحنى لا تمثل احتمالاً وإنما تمثل قيمة ما تسمى بدالة كثافة الاحتمال " Probability density function" عند هذه النقطة وتناظر دالة الاحتمال في المتغيرات العشوائية المنفصلة. والاحتمال في التوزيعات المتصلة هو المسافة الموجودة تحت منحنى دالة كثافة الاحتمال المناظرة لفترة ما، وليس لنقطة ما .

نلاحظ في الشكل (4.4) أن مساحة المنطقة المظللة المحصورة بين b , a تساوي احتمال أن يأخذ المتغير العشوائي المتصل X أية قيمة في الفترة من a إلى b أي أنها تساوي $P(a < x < b)$ والجدير بالملاحظة أن المساحة الكلية تحت منحنى دالة كثافة الاحتمال لأي متغير عشوائي متصل تساوي الواحد الصحيح.

شكل (4.4)

يوضح رسم دالة الكثافة الاحتمالية $P(a < x < b)$



وتأخذ دالة كثافة الاحتمالي للتوزيع الطبيعي الشكل التالي:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad , \quad \begin{array}{l} -\infty < x < +\infty \\ -\infty < \mu < +\infty \\ 0 < \sigma < +\infty \end{array}$$

حيث :

π : مقدار ثابت يساوي تقريباً 3.1416

e: قيمة ثابتة (ثابت تنبيري) تساوي2.7183 تقريباً وهي الأساس للوغاريتم الطبيعي

μ : الوسط الحسابي لقيم المتغير العشوائي X

σ : الانحراف المعياري

f(x): ارتفاع المحور العمودي (التكرار) للمنحنى

وتعتمد هذه المعادلة على معلمتي الوسط الحسابي μ والانحراف المعياري σ

1.2.3 خصائص منحنى التوزيع الطبيعي :

1- شكله متماثل ويشبه الجرس مرتفع من الوسط ومنخفض بشكل تدريجي عند الطرفين ولكونه متماثل فإن 50٪ من مساحته تقع إلى يمين المتوسط الحسابي و 50٪ على يسار المتوسط الحسابي μ.

2. يمتد طرفا المنحنى من $-\infty$ إلى $+\infty$ ويكون عدد التكرارات صغيرة عند طرفي المنحنى ولا يلتقيان بالمحور السيني أبداً وتزيد كلما اقتربت من مركز المنحنى.

3- المتوسطات تكون متساوية أي أن الوسط الحسابي= الوسيط = المنوال

4- المجموع الكلي للمساحة تحت منحنى التوزيع الطبيعي تساوي الواحد الصحيح.

5- تتوزع المساحة تحت المنحنى على المحور الأفقي (العدد الكلي لمفردات العينة المعيارية) كالأتي

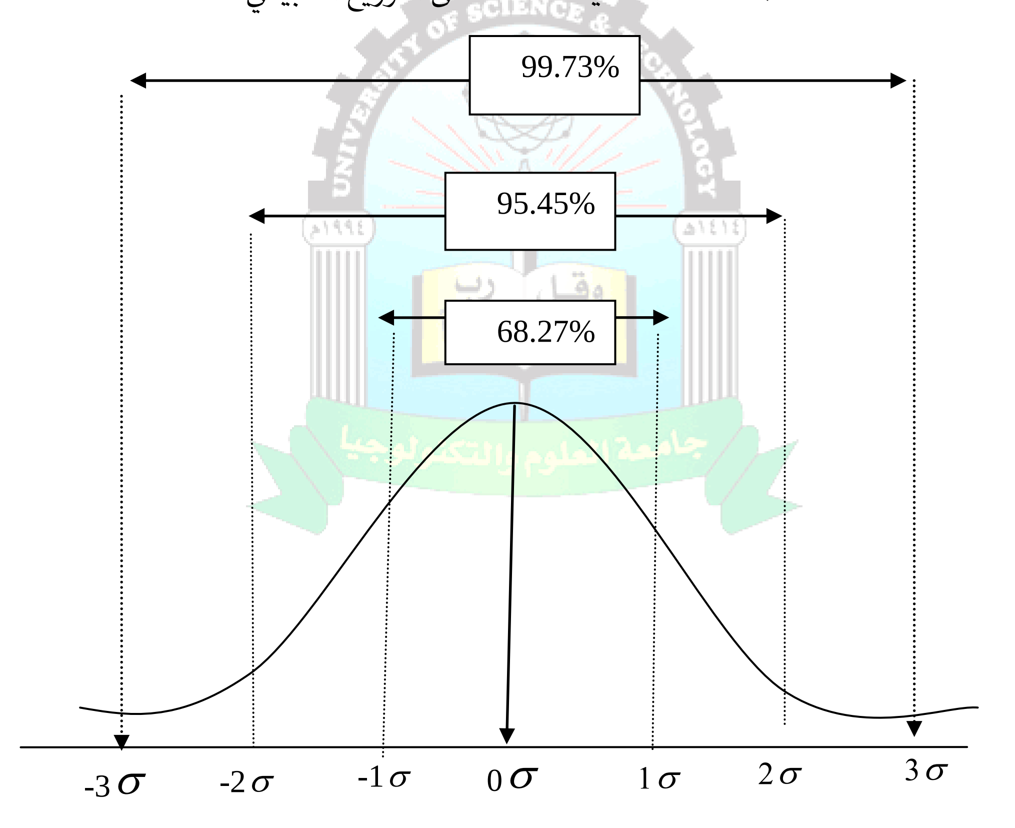
68.27٪ تقريباً من مساحة الشكل الموزع طبيعياً تقع بين $\mu \pm \sigma$

95.45٪ تقريباً من مساحة الشكل الموزع طبيعياً تقع بين $\mu \pm 2\sigma$

99.73٪ تقريباً من مساحة الشكل الموزع طبيعياً تقع بين $\mu \pm 3\sigma$

شكل (5.4)

يمثل نسب المساحة النسب المختلفة للبيانات الموزعة طبيعياً بالنسبة للمساحات الكلية تحت منحنى التوزيع الطبيعي



6- يعتمد شكل المنحنى اعتماداً كاملاً على معلمتي الوسط الحسابي μ والانحراف المعياري σ .

7- التوقع = μ ، التباين = σ^2 ، الانحراف المعياري= σ

2.2.3 التوزيع الطبيعي القياسي (Standard Normal Distribution):

عرفت **عزيزي الدارس** في الفقرة السابقة أن التوزيع الطبيعي يعتمد على معلمتي الوسط الحسابي والانحراف المعياري وأن المجموع الكلي للمساحة تحت منحنى التوزيع الطبيعي تساوي الواحد الصحيح. وقد نحتاج إلى حساب جزء من هذه المساحة والتي تنحصر عادة بأصغر من قيمة ما أو أكبر من قيمة ما، أو المساحة المحصورة بين قيمتين. ولحساب المساحة بين قيمتين ولتكن (X_1 , X_2) فإنه يمكن الاستعانة بجداول خاصة بحساب المساحات للتوزيع الطبيعي، إلا أن مثل هذه الجداول سيكون عددها لا نهائي طالما أن متوسط المجتمع وانحرافه المعياري يأخذان قيماً كثيرة جداً وأن قيم المتغير العشوائي X لا نهائية أيضاً. لذا تم الاستعانة بالجزء الخاص من دالة التوزيع الطبيعي والذي يمثل قوة العدد الأسي وهو $\frac{X-\mu_x}{\sigma_x}$ والذي يمثل القيمة المعيارية Z التي تمت دراستها في الوحدة الثانية من هذا المقرر، والتي متوسطها (القيمة المعيارية) صفر وتباينها يساوي الواحد الصحيح.

لذا يمكن تحويل قيم المتغير العشوائي المعتدل إلى قيم معيارية أي تحويله إلى متغير عشوائي جديد يسمى بالتوزيع الطبيعي القياسي ويرمز له بالرمز Z (وهو التوزيع الطبيعي الذي وسطه صفر وتباينه واحد) أي أن:

$$Z = \frac{x-\mu}{\sigma}$$

حيث X : متغير عشوائي طبيعي

μ : الوسط الحسابي للمتغير x

σ : الانحراف المعياري للمتغير x

Z : تسمى بالقيمة المعيارية، أما القيمة الاحتمالية لـ Z فيتم إيجادها باستخدام الجداول الإحصائية للتوزيع الطبيعي القياسي.

1.2.2.3 استخدام جدول المساحات تحت المنحنى الطبيعي المعياري:

نظراً لتماثل نصفي المنحنى الطبيعي فإن جدول مساحة المنحنى الطبيعي يُظهر عادة نسبة المساحة في النصف الأيمن من المنحنى وهو الجزء الموجب أي الذي يقابل قيمة موجبة للمتغير بعد تحويله إلى قيمة معيارية، وفي الوقت نفسه تكون قيمة (Z) السالبة مطابقة للقيمة الموجبة إلا أنها تقع في النصف الأيسر من المنحنى، وبما أن المساحة تحت المنحنى تساوي واحد صحيح فإن المساحة على يمين الوسط الحسابي للمنحنى تساوي 0.5 وهي المساحة التي تظهر في الجداول الخاصة بالمساحة والتي تبدأ القياس من المنتصف متجهة نحو اليمين حيث تكون المساحة عند الوسط الحسابي تساوي صفراً وتزداد المساحة كلما اتجهنا نحو اليمين أي بزيادة قيمة (Z).

ولتقدير قيمة المتغير (Z) طبقاً لهذا الجدول تستخدم قيمة (Z) المطلوبة من الجدول مباشرة وتكون هي قيمة التكرار النسبي أو احتمال أن تقع القيمة بين الوسط الحسابي وقيمة (Z) ملحق رقم (1). ولا شك أن هذا الجدول يسهل طريقة الحصول على الاحتمالات بدلالة القيمة المعيارية (Z). والأمثلة الآتية توضح ذلك:



مثال (27.4):

إذا كان لديك المتغير العشوائي X يتبع التوزيع الطبيعي بمتوسط 40 وانحراف معياري 5

المطلوب: إيجاد القيم المعيارية المقابلة للقيم $x_1=45, x_2=35$

الحل:

$\because X \sim N(40\,,\,5)$ فإن المتغير العشوائي الذي توزيعه التوزيع الطبيعي المعياري هو

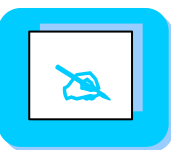
$$\because\ \sigma = 5, \bar{x} = 40$$

$$\therefore\ z = \frac{x_i - \bar{x}}{\sigma}$$

$$z_1 = \frac{45-40}{5} = \frac{5}{5} = 1$$

$$z_2 = \frac{35-40}{5} = \frac{-5}{5} = -1$$

وهذه تسمى بالقيم المعيارية كما مر معنا سابقاً.



مثال(28.4):

إذا كانت درجات مجموعة مكونة من 500 شخص تقدموا للعمل بإحدى المؤسسات تتبع التوزيع الطبيعي بوسط حسابي قدره 70 درجة وانحراف معياري قدرة 5 درجات فاحسب عدد الأشخاص فيما يلي:

1. الحاصلون على 66 إلى 76 درجة $P(66 \le x \le 76)$
2. الحاصلون على أكثر من 78 درجة $P(x > 78)$
3. الحاصلون على أقل من 60 درجة $P(x < 60)$

الحل:

1. حتى تتمكن من حساب كل الاحتمالات السابقة يجب أولاً تحويل بُعد القيمة عن الوسط الحسابي إلى قيمة معيارية في كل حالة أي: $Z = \frac{X_i - \bar{X}}{\sigma}$

$$1. \quad Z_1 = \frac{x - \bar{x}}{\sigma} = \frac{66 - 70}{5} = \frac{-4}{5} = -.8$$

$$2. \quad Z_2 = \frac{x - \bar{x}}{\sigma} = \frac{76 - 70}{5} = \frac{6}{5} = 1.2$$

نلاحظ أن المطلوب هو حساب احتمال أن تكون قيمة X أكبر من 66 وأقل من 76 وبالتالي المطلوب هو مجموع مساحة منطقتين حول الوسط الحسابي فيجب معرفة مساحة كل منطقة ومن ثم جمعهما.

∴ المنطقة الأولى تبعد انحراف معياري قدره 0.8 إلى يسار الوسط الحسابي وبالتالي تكون المساحة (الاحتمال) من ملحق (1) هو:

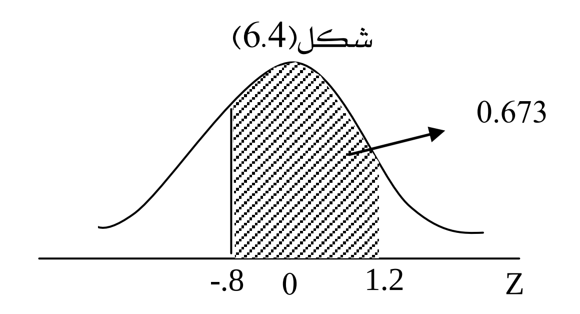
$$P(-.8 \le Z \le 0) = P(0 \le Z \le .8) = 0.2881$$

وتبعد النقطة الثانية 1.2 انحراف معياري إلى يمين الوسط الحسابي وبالتالي تكون المساحة (الاحتمال) هي: $P(0 \le Z \le 1.2) = 0.3849$

وبذلك تكون المساحة المطلوبة هي:

$$P(-.8 \le z \le 1.2) = 0.2881 + 0.3849 = 0.6730$$

شكل(6.4)



فيكون عدد الأشخاص الحاصلون على 66 إلى 76 درجة هو:

$0.6730 * 500 \approx 337$ شخص

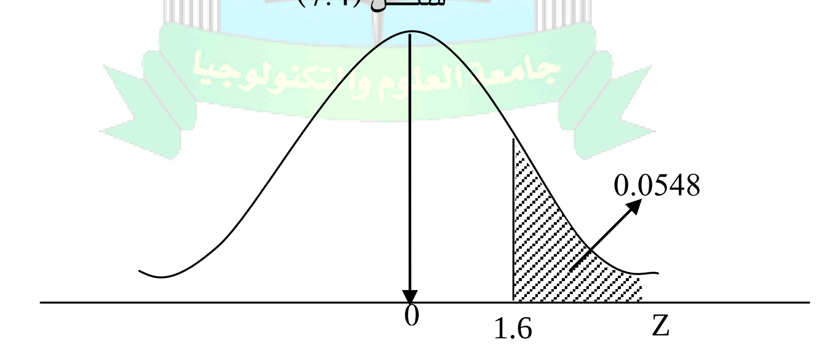
$$2- \quad Z = \frac{x_i - \bar{x}}{s} = \frac{78 - 70}{5} = 1.6$$

هذه القيمة تبعد 1.6 انحراف معياري على يمين الوسط الحسابي وهي تحصر مساحة قيمتها $P(z > 1.6) = 0.4452$ بين الوسط الحسابي ونقطة تبعد 1.6 انحراف معياري، ولكن المطلوب هو المساحة إلى يمين هذه النقطة وبالتالي تكون المساحة المطلوبة هي:

$$P(z > 1.6) = P(Z > 0) - P(0 < z < 1.6) = 0.5 - 0.4452$$

$$P(z > 1.6) = 0.0548$$

شكل (7.4)



فيكون عدد الأشخاص الحاصلون على أكثر من 78 درجة هو شخص

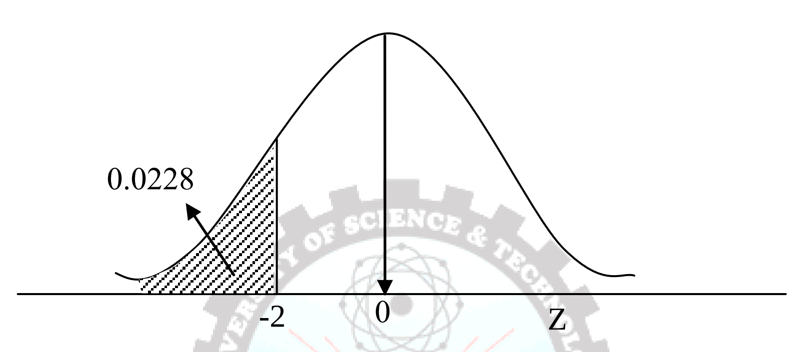
$0.0548 * 500 \approx 27$

$$3- \quad Z = \frac{x_i - \bar{x}}{\sigma} = \frac{60 - 70}{5} = -2$$

فإن هذه القيمة تبعد انحرافين معياريين على يسار الوسط الحسابي والمطلوب هو احتمال أن تكون x أقل من هذه القيمة وبالتالي هذه النقطة تحصر مساحة مقدارها

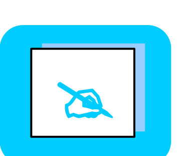
$$P(X < 60) = P(Z < -2) = P(-2 < Z < 0)$$
$$= P(Z < 0) - P(0 < Z < 2)$$
$$= 0.5 - 0.4772 = 0.0228$$

شكل(8.4)



ويكون عدد الأشخاص الحاصلون على أقل من 60 درجة هو: شخص

$$0.0228 * 500 \approx 11$$



مثال(29.4):

أوجد احتمال أن Z :

1- أقل من 1.64

2- أكبر من 1.64

الحل:

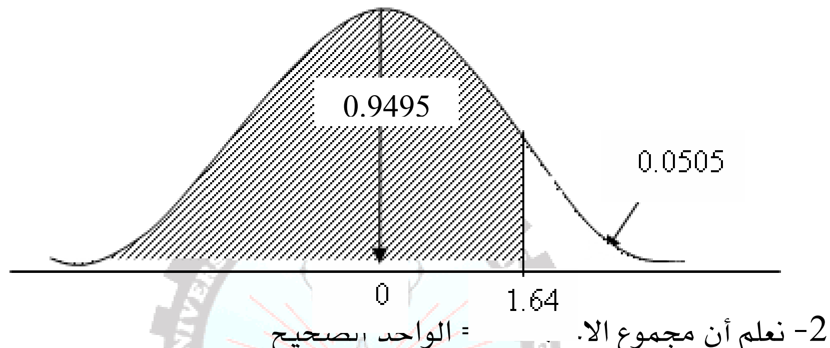
لإيجاد هذه الاحتمالات نستعين بالملحق(1) وبرسم المنحنى الاعتدالي المعياري كما يلي :

1. $P(Z < 1.64)$ من جدول ملحق (1) نحصل على القيمة المعيارية Z بجمع القيمتين المجودتين بالصف العلوي والعمود الأول يسار الجدول . ويحتوي العمود الأول من جهة اليسار على قيم تصل إلى رقم عشري واحد فقط بينما يحتوي الصف العلوي على الرقم المئوي .

فالاحتمال المتجمع المناظر للقيمة 1.64 يوجد أمام الصف 1.6 وتحت العمود 0.04 (لا حظ أن 1.6 + 0.04 = 1.64 وهي قيمة Z المطلوب أيجاد الاحتمال المتجمع عندها) وهذا الاحتمال هو0.4495 أي أن $P(Z < 1.64) = 0.4495$ وهذا هو الاحتمال المتجمع للمتغير Z من $-\infty$ إلى 1.64 والمساحة المظللة في شكل (9.4) تمثل هذا الاحتمال.

شكل (9.4)

يوضح المساحة تحت المنحنى الطبيعي القياسي على يمين ويسار 1.64+

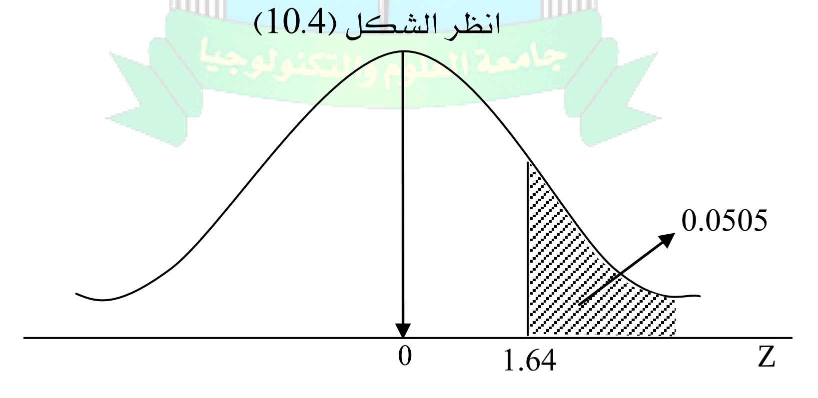


2- نعلم أن مجموع الا… = الواحد الصحيح

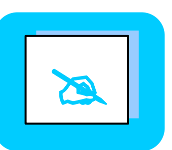
وحيث أن المساحة الكلية تحت منحنى أي متغير عشوائي متصل تمثل 0.5 على يمين الوسط الحسابي و0.5 على يساره لذا فإن:

$$P(Z > 1.64) = 0.5 - P(Z < 1.64)$$
$$= 0.5 - 0.4495 = 0.0505$$

انظر الشكل (10.4)



مثال (30.4):



إذا كـان X يمثـل توزيعـاً احتماليـاً طبيعيـاً بمتوسـط 40 وانحـراف معياري 5 فأوجد الاحتمالات الآتية:

$$1-P(40\le X\le 45)$$

$$2-P(50\le X\le 55)$$

$$3-P(35\le X\le 40)$$

الحل:

نحول قيم X للتوزيع الطبيعي إلى قيم Z المعيارية باستخدام الصيغة

$$Z=\frac{X-\mu}{\sigma}$$

$$1-P(40\le X\le 45)$$

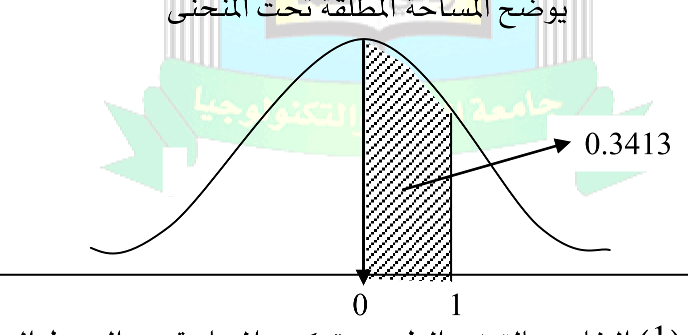
$$Z_1=\frac{40-40}{5}=0$$

$$Z_2=\frac{45-40}{5}=1$$

$$P(40\le x\le 45)=P(O\le Z\le 1)$$

شكل (11.4)

يوضح المساحة المطلقة تحت المنحنى



ومن الملحق (1) الخاص بالتوزيع الطبيعي تكون المساحة بين الوسط الحسابي ونقطة تبعد انحراف معياري واحد هي:

$$P(O\le Z\le 1)=0.3413$$

$$2.\quad P(50\le X\le 55)$$

$$Z_1=\frac{50-40}{5}=2$$

$$Z_2 = \frac{55-40}{5} = 3$$

$$P(50 \le X \le 55) = P(2 \le Z \le 3)$$

ومعنى ذلك أننا نريد إيجاد المساحة المحصورة ما بين $(0, 3)$ وهي:

$$p(0 \le Z \le 3) = 0.4987$$

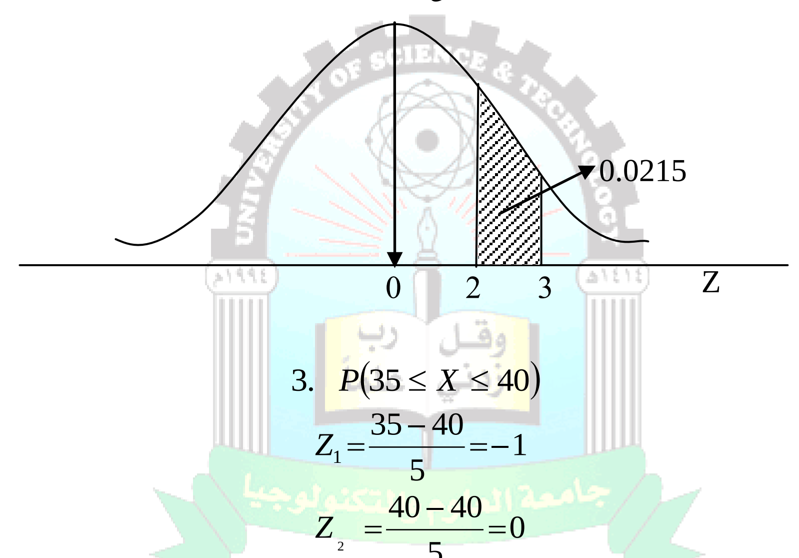
وكذلك المساحة المحصورة بين $(0,2)$ وهي:

$$p(0 \le Z \le 2) = 0.4772$$

وبطرح المساحة الثانية من الأولى نحصل على المساحة المطلوبة وهي:

$$p(2 \le Z \le 3) = 0.4987 - 0.4772 = 0.0215$$

شكل (12.4)



3. $P(35 \le X \le 40)$

$$Z_1 = \frac{35-40}{5} = -1$$

$$Z_2 = \frac{40-40}{5} = 0$$

$$\therefore P(35 \le X \le 40) = p(-1 \le Z \le O)$$

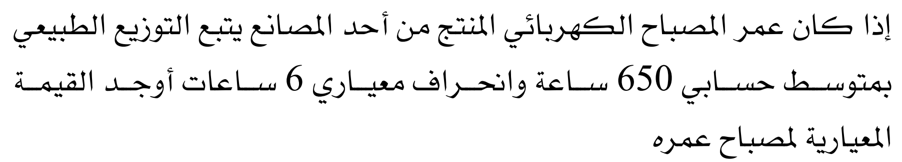
$$= P(0 \le Z \le 1) = 0.3413$$

تدريب (6)



1. عرف المتغير العشوائي واذكر أقسامه ؟
2. أذكر بعض خصائص التوزيع الطبيعي ؟

تدريب (7)

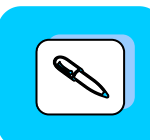




إذا كان عمر المصباح الكهربائي المنتج من أحد المصانع يتبع التوزيع الطبيعي بمتوسط حسابي 650 ساعة وانحراف معياري 6 ساعات أوجد القيمة المعيارية لمصباح عمره

1. 700 ساعة
2. 690 ساعة

تدريب (8)



معروف بأن أطوال الأنابيب التي ينتجها مصنع ما تقريباً تتوزع طبيعياً بوسط 20 Cm وانحراف معياري 3 Cm أكدت إحدى الطلبيات ضرورة أن تكون على الأقل 95٪ من الأنابيب أطول من 20 Cm فهل يحقق منتج هذا المصنع هذا الطلب ؟

أسئلة التقويم الذاتي (3) :



1-عرف كلاً من :

1) المتغير العشوائي .　　　　2) قانون التوزيع الاحتمالي.

2- إذا كان الوسط الحسابي لأطوال ألف أنبوبة حديد يساوي 80 سم والانحراف المعياري 15 فأوجد

أ- نسبة وعدد الأنابيب التي تقع أطوالها بين 80- 120 سم ؟

ب- نسبة وعدد الأنابيب التي تقع أطوالها بين 85-125 سم ؟

ج- نسبة وعدد الأنابيب التي يقل طولها عن 65.4 سم ؟

3- إذا بلغ عدد طلاب المستوى الرابع محاسبة 300 طالب وكان متوسط أوزانهم 60 كغم والانحراف المعياري 7 . فما هي نسبة من تقع أوزانهم بين 50-65 كغم ؟ وكم عددهم ؟

4. اختبار الفروض (Hypothesis Testing):

عزيزي الدارس، إن اختبار الفرضيات يعتبر أحد المواضيع الرئيسة للاستدلال الإحصائي ويهدف للوصول إلى قرار بشأن معلمة المجتمع من خلال قبول أو رفض تقديرها المعتمد على معطيات العينة المسحوبة من ذلك المجتمع.

ويعرف اختبار الفروض بأنه قاعدة أو إجراء تؤدي إلى رفض أو عدم رفض فرض العدم . ويتم اختبار الفروض بتقسيم فضاء العينة لكل النتائج الممكنة للتجربة العشوائية إلى قسمين غير متداخلين-(أحدهما النتائج التي إذا حدثت لا نرفض فرض العدم والآخر للنتائج التي إذا حدثت نرفضه)- ويسمى هذان القسمان منطقة القبول "acceptance region" ومنطقة الرفض "rejection region" على الترتيب. ويطلق على القيمة التي تفصل بين منطقة القبول ومنطقة الرفض القيمة (الحرجة) **ويكون القرار برفض فرض العدم بينما هو في الواقع صحيح، وبذلك نقع في خطأ من النوع الأول " type 1 error " كذلك قد يحدث عدم رفض فرض العدم وهو في الحقيقة خطأ ونسمي هذا خطأ من النوع الثاني " type 11 error "**

فمثلاً: أفترض أن أحد مصانع أقلام الرصاص يشتري المحايات من مورد معين وأن المسئول عن المصنع يجب أن يقرر قبول أو رفض أية شحنة تصل إلى المصنع من هذا المورد. ولقد قرر المسئول قبول أية شحنة إذا كانت نسبة المعيب فيها أقل من أو يساوي 15٪ وبالطبع فإنه لا يستطيع أن يفحص كل ممحايات الشحنة . لذا قرر أن يسحب عينة عشوائية من كل شحنة حجمها 20 ممحاة وتقبل الشحنة إذا وجد بالعينة ثلاث ممحايات أو أقل معيبة أما إذا وجد في العينة أكثر من ثلاث ممحايات معيبة فإن الشحنة سترفض بكاملها.

ومن الواضح أن بيانات العينة قد تؤدي إلى قبول الشحنة بينما تكون نسبة المعيب الفعلية فيها أكبر من 15٪ وهذا قرار خاطئ، والعكس يمكن أن تؤدي بيانات العينة إلى رفض الشحنة بينما تكون نسبة المعيب الفعلية فيها أقل من أو تساوي 15٪ وهذا قرار خاطئ .

وعموماً فإن مدير المصنع يستخدم بيانات العينة للإجابة على السؤال التالي "هل نسبة المعيب في الشحنة كبيرة بدرجة توجب رفضها " ؟ وللإجابة على هذا السؤال فإن مدير المصنع يكون قد اتخذ قراراً بالنسبة لمعلمة المجتمع (نسبة المعيب

في الشحنة). فإذا عرفنا أن اختبارات الفروض تؤدي في النهاية إلى اتخاذ قرار معين نكون قد عرفنا أن المدير قد حقق الهدف من اختبار الفرضيات عند إصداره القرار الخاص بقبول أو رفض الشحنة.

صياغة الفرضية الإحصائية (Statistical Hypothesis)

الفرضية الإحصائية: هي عبارة عن ادعاء أو افتراض أو تصريح (قد يكون صواباً أو خطأً) حول متغير عشوائي أو متغيرات عشوائية.

أو هو: عبارة عن صياغة مبدئية حول معلمة أو أكثر من معالم المجتمع المجهولة أو لمجموعة من المجتمعات.

والوظيفة الأساسية لاختبار الفروض الإحصائية في ظل عدم التأكد من طبيعتها هي مساعدتنا في اتخاذ قرارات رشيدة بخصوص هذه المعالم . ونتيجة لذلك فإن الإحصاء الحديث يشار إليه غالباً بأنه (**علم اتخاذ القرار في ظل عدم التأكد**) واتخاذ القرار ما هو إلا اختيار من بين بدائل.

1.4. الخطوات الأساسية في اختبارات الفروض:

تعد اختبارات الفروض الإحصائية أسلوباً لتحديد ما إذا كانت بيانات العينة تؤدي إلى قبول أو رفض صياغة مبدئية عن معلمة (أو معالم) المجتمع وتتضمن الخطوات الآتية:

1- التعرف على توزيع المجتمع الإحصائي.

2- صياغة فرض العدم (null hypothesis).

والفرض البديل (alternative hypothesis).

3- تحديد مستوى المعنوية (level of Significance).

4- استخراج القيمة الجدولية.

5- صياغة قاعدة القرار.

6- اتخاذ القرار الإحصائي. وتنتهي هذه العملية باتخاذ قرار برفض أو عدم رفض فرض العدم.

ويمكن تقسيم الفرضيات إلى قسمين هما:

أ-فرض العدم (الفرضية الصفرية) "null hypothesis "

هو عبارة عن صياغة مبدئية عن معلمة المجتمع المجهولة ويرمز له بالرمز H_O. وتوحي هذه الصيغة إلى فكرة عدم وجود فرق بين معلمة المجتمع وقيمة معينة وأن هذا الفرق المشاهد هو فرق ظاهري مرده للأخطاء العشوائية لذا تسمى هذه الصيغة بالفرضية الصفرية أو فرض العدم.وهي على ثلاثة أنواع، فعلى سبيل المثال إذا كان الاختبار يتعلق بمتوسط المجتمع μ_X ، فإن فرضية العدم ستكون واحدة من بين الصيغ الآتية:

1. في حالة مساواة معلمة لقيمة افتراضية. $H_0 : \mu_X = \mu_O$
2. في حالة كون المعلمة أقل من أو تساوي قيمة افتراضية. $H_0 : \mu_X \le \mu_O$
3. في حالة كون المعلمة أكبر من أو تساوي قيمة افتراضية . $H_0 : \mu_X \ge \mu_O$

حيث μ_O هي قيمة رقمية.

فإذا أردنا معرفة ما إذا كان متوسط درجات الذكاء للطلاب في المرحلة الأساسية بالجمهورية اليمنية يختلف عن 100 درجة، فإننا نفترض أن المتوسط لا يختلف عن 100 درجة وبالتالي يصاغ فرض العدم حسب الصيغ:

$$H_0 : \mu = 100$$

$$H_0 : \mu \le 100$$

$$H_0 : \mu \ge 100$$

ب-الفرض البديل (Alternative Hypothesis)

هو عبارة عن صياغة مبدئية تشير إلى أن نفس المعلمة المجهولة لها قيمة تختلف عن القيمة التي حددها فرض العدم ويرمز له بالرمز H_1 ، وهي على ثلاثة أنواع، فعلى سبيل المثال إذا كان الاختبار يتعلق بمتوسط المجتمع μ_X ، فإن الغرض البديل سيكون واحد من بين الصيغ الآتية:

1. في حالة عدم مساواة معلمة لقيمة افتراضية. $H_0 : \mu_X \ne \mu_O$
2. في حالة كون المعلمة أقل من أو تساوي قيمة افتراضية. $H_0 : \mu_X < \mu_O$
3. في حالة كون المعلمة أكبر من أو تساوي قيمة افتراضية . $H_0 : \mu_X > \mu_O$

ويصاغ الفرض البديل للمثال السابق على إحدى الصور الآتية:

$$H_0 : \mu \neq 100$$
$$H_0 : \mu < 100$$
$$H_0 : \mu > 100$$

2.4 الأخطاء (Errors)

1. **الخطأ من النوع الأول:** يحدث الخطأ من النوع الأول عند رفض فرض العدم H_o بينما في الحقيقة يجب قبوله لأن هذا الرفض ناتج عن خطأ في معطيات العينة، واحتمال الوقوع في مثل هذا النوع من الخطأ يرمز له بالرمز $\propto$ (ألفا).

2. **الخطأ من النوع الثاني:** يقع الخطأ من النوع الثاني في حالة قبولنا لفرضية العدم H_o بينما هي في الحقيقة خطائة، واحتمال الوقوع في مثل هذا النوع من الخطأ يرمز له بالرمز β (بيتا) وكلما قلت قيمة $\propto$ والتي ترمز لمستوى معنوية الاختبار كلما قل احتمال الوقوع في الخطأ من النوع الأول. والجدول الآتي يوضح هذين النوعين من الأخطاء:

القرار	نوع الخطأ	احتمال الخطأ
رفض H_o وهو صحيح	I	$\propto$
قبول H_o وهو خاطئ	II	β

ولتقريب صورة وقوع هذه الأخطاء، نفرض بأن متوسط وزن الرجل في المجتمع هو أقل من 65 كيلوغرام فتكون فرضية العدم هي:

فرضية العدم هي: $H_0 : \mu < 65$

والفرضية البديلة هي: $H_1 : \mu > 65$

وبما أن تقليل احتمال الوقوع في الخطأ من النوع الأول يتم من خلال رفع مستوى المعنوية التي تعتمد للتمييز إن كان الخطأ معنوي أم غير معنوي، نفرضها مثلاً 0.05 بدلاً من 0.01، إلا أن ذلك يؤدي إلى زيادة احتمال الوقوع في خطأ من النوع الثاني، كما أن تقليل الخطأ من النوع الثاني يؤدي إلى زيادة احتمال الوقوع في الخطأ من النوع الأول . وعليه فإنه عند دراسة الموضوع يؤخذ بعين الاعتبار أي من الخطأين هو الأخطر . فمثلاً: إذا كان الاختبار يتم لعينة من اسطوانات الغاز ذات الاستخدام المنزلي أو الصناعي والاختبار يتم على سماكة المعدن المستخدم في صنع

هذه الاسطوانات فالوقوع بالخطأ من النوع الأول هنا هو رفض الاسطوانات وهي تحقق المواصفات المطلوبة، وبالتالي يعاد تصنيعها.

أما الوقوع بخطأ من النوع الثاني فهو قبول الاسطوانات وهي لا تحقق المواصفات المطلوبة وبالتالي لا يمكن تلافي الآثار الناجمة عن هذا الخطأ حيث من الممكن أن تكون هناك خسائر مادية أو بشرية أو الاثنين معاً نتيجة لتفجر الاسطوانات أثناء الاستخدام، وفي مثل هذه الحالة يجب أن يكون مستوى المعنوية كبيراً لتقليل احتمال الوقوع بخطأ من النوع الثاني.

ومثال أخر : نفرض أن شركة أدوية أدعت أنها اكتشفت علاجاً لمرض الإيدز، فالوقوع بخطأ من النوع الثاني هو قبول الدواء رغم عدم صحة هذا الادعاء وإعطائه للمرضى اليائسين، أما الوقوع بخطأ من النوع الأول فهو رفض هذا الدواء وبالتالي حرمان المرضي من فرصة العلاج والتي يمكن أن تكون موجودة . لذا في مثل هذه الحالة يجب تقليل احتمال الوقوع بخطأ من النوع الأول لتكبير احتمال الوقوع بخطأ من النوع الثاني إما بواسطة زيادة حجم العينة أو رفع مستوى المعنوية أو زيادتهما معاً.

وتوجد أربعة قرارات ممكنة يمكن اتخاذها في أي مشكلة من مشاكل اختبارات الفروض هي حسب الجدول الآتي:

القرار	H_0 صحيح	H_0 خاطئ
رفض H_o	خطأ من النوع الأول ($\propto$)	قرار سليم $(1-\beta)$
عدم رفض H_o	قرار سليم $(1-\propto)$	خطأ من النوع الثاني (β)

3.4 مستوى المعنوية (Level of Significance):

إن مستوى المعنوية هو عبارة عن " احتمال رفض فرض العدم وهو صحيح " أي احتمال الوقوع في خطأ من النوع الأول Type 1 error ويرمز له بالرمز α

وبناء عليها تتحدد منطقة أو مساحة الرفض تحت منحنى توزيع إحصاءه الاختبار، وتؤثر قيمة α على القرار الخاص، باعتبار الفرق بين القيمة التي نحصل عليها من العينة والقيمة التي يحددها فرض العدم فرقاً معنوياً، باعتباره فرقاً لا يمكن إرجاعه للصدفة.

4.4 خطوات إجراء الاختبار:

1- تحديد فرضية العدم H_0 التي تنص على أنه لا يوجد فرق معنوي بين الإحصاءه وبين المعلمة وأن هذا الفرق هو فرق ظاهري يعود للأخطاء العشوائية.

2- تحديد الفرضية البديلة H_1 .

وقد يكون الاختبار من جهة واحدة أو من جهتين ويقصد بذلك أن الانحراف عن فرضية العدم قد يكون باتجاه واحد أو أنها موزعة على اتجاهين وهذا يعتمد على صيغة فرضية العدم . فإذا كانت محددة مسبقاً بأكبر أو أقل أي:

$$H_0 : \mu \geq \mu_0$$

أو

$$H_0 : \mu \leq \mu_0$$

فهذا معناه أن الاختبار من جانب واحد ، ففي حالة رفض فرضية العدم فإن الفرضية البديلة هي H_1 يكون اتجاهها معلوماً .

$$H_1 : \mu < \mu_0$$

أو

$$H_1 : \mu > \mu_0$$

أما في حالة فرضية العدم على الصورة: $H_0 : \mu = \mu_0$

فإن المتوقع في حالة رفض هذه الفرضية أن تكون الفرضية البديلة على الصورة:

$$H_0 : \mu_X \neq \mu_O$$

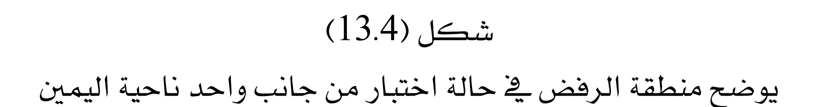
$$H_0 : \mu_X < \mu_O$$

$$H_0 : \mu_X > \mu_O$$

وهذا يعني عدم معلومية الاتجاه الذي ستكون عليه مسبقاً وبهذا تتوزع على جانبي التوزيع فمثلاً : إذا فرضنا أن إحدى شركات النقل الجوي تنوي تخفيض أجور النقل فمن المؤكد ارتفاع الطلب أي: $H_0 : \mu \geq \mu_0$ أما إذا أرادت تغيير شعار الطائرة وهي غير متأكدة أن هذا التغيير سوف يؤدي إلى تراجع الإقبال عليها فإن:

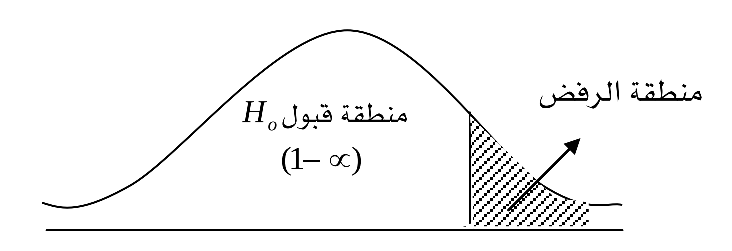
$$H_0 : \mu = \mu_0$$

ويكون رفض الفرض H_0 أما بسبب $\mu > \mu_0$ أو $\mu < \mu_0$ ويصبح الاختبار من جانبين والأشكال (13.3) و (14.3) و(15.3) توضح حالة الاختبار من جانب واحد ومن جانبين.



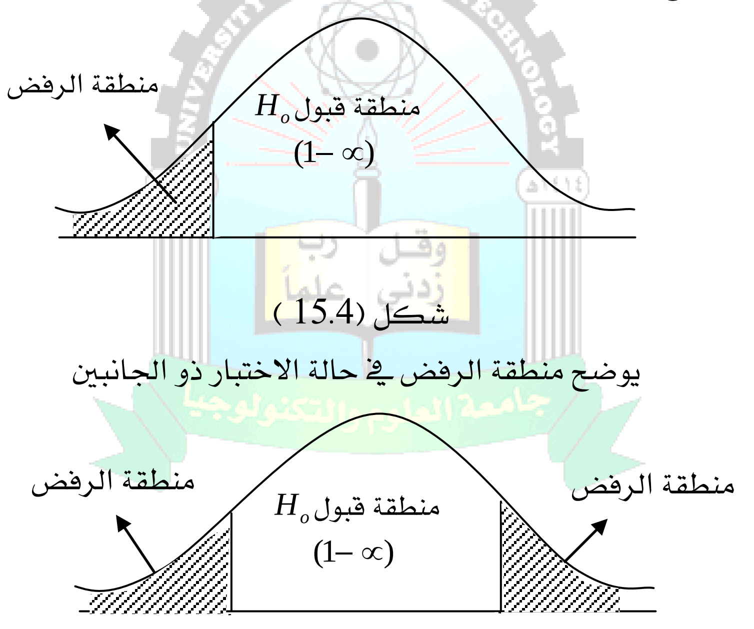
شكل (13.4)

يوضح منطقة الرفض في حالة اختبار من جانب واحد ناحية اليمين



شكل (14.4)

يوضح منطقة الرفض في حالة اختبار من جانب واحد ناحية اليسار



شكل (15.4)

يوضح منطقة الرفض في حالة الاختبار ذو الجانبين

5.4 قوة الاختبار (Power of The Test):

ذكرنا سابقاً أنه في حالة قبولنا لفرض العدم H_o وهو غير صحيح يؤدي إلى وقوعنا في خطأ من النوع الثاني، وأن احتمال الوقوع في هذا النوع من الخطأ يدعى β (قوة الاختبار) وعلى هذا الأساس فإن هذا الاحتمال يعتمد على حجم العينة n

والانحراف المعياري للمجتمع وعلى مستوى المعنوية α ونوع الاختبار إن كان من جانب واحد أو من جانبين وتحسب من العلاقة:

$$\beta = \sqrt{n}\,(\mu - \mu_0)\,\sigma$$

والجدير بالملاحظة أنه إذا كان حجم العينة كبيراً ($n \geq 30$) فتوزيع المعاينة هو التوزيع الطبيعي، وإذا كان الاختبار يهدف لمقارنة الوسط الحسابي $\overline{X}$ للعينة مع الوسط الحسابي للمجتمع μ فإن الخطأ المعياري يحسب من العلاقة :

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

وأن الإحصاءه Z تمثل المنطقة الحرجة من خلال تحويل القيم العشوائية X_i إلى قيم طبيعية معيارية Z_i باستخدام الصيغة :

$$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

أما إذا كان حجم العينة صغيراً ($n < 30$) فتوزيع المعاينة هو توزيع (ستودنت) (إحصاءة t) بدرجة حرية $n - 1$ وصيغته الإحصائية هي:

$$t = \frac{\overline{X} - \mu}{S/\sqrt{n}}$$

حيث S هي الانحراف المعياري



مثال (31.4):

أراد أحد البنوك التأكيد من أن فترة انتظار العميل لا تزيد عن 20 دقيقة $\mu \leq 20$ وبانحراف معياري $\sigma = 10$ دقيقة . أخذت عينة عشوائية حجمها 100 عميل وسجلت فترة الانتظار لكل منهم متوسطها $\bar{x} = 21$ دقيقة . المطلوب هو اختبار فرضية البنك عند مستوى معنوية $\propto = 0.05$

الحل:

1. صياغة الفرضيات

$$\because H_o : \mu \leq 20$$
$$H_1 : \mu > 20$$

2. تحديد مستوى المعنوية ومساحة مناطق القبول والرفض والنقاط الحرجة:

مستوى المعنوية $\propto = 0.05$ وفرضية الاختبار من جهة واحدة، مما يشير إلى منطقة واحدة لرفض H_o ستكون من جهة يمين ذيل المنحنى. ولحساب مساحتها نقوم بحساب منطقة القبول حيث:

$$\because 1-\propto = 1 - 0.05 = 0.9500$$

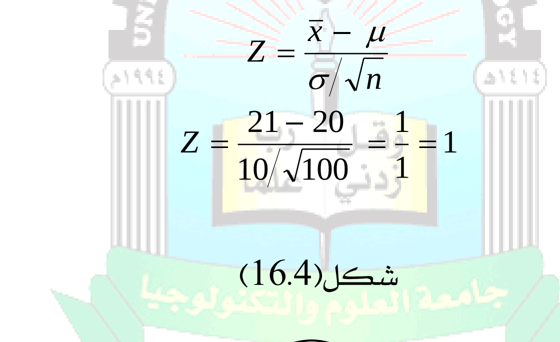
وحيث أن الجدول المتاح في الملحق رقم (1) يعرض نصف المساحة تحت منحنى التوزيع الطبيعي أي 0.5، وعليه فإن:

$$0.9500 - 0.5000 = 0.4500$$

وهذا الاحتمال " 0.4500 " غير موجود صراحة في ملحق رقم (1) والملاحظ أن أقرب رقم له هو : 0.4505 وهذا الاحتمال يقابل قيمة Z التي تساوي 1.65

أي أن المساحة الأقل من القيمة المعيارية 1.65هي 0.4505، أنظر ملحق (1) وهنا تكون مساحة رفض H_o هي 0.05 لذا تكون القيمة المعيارية 1.65 هي النقطة الحرجة

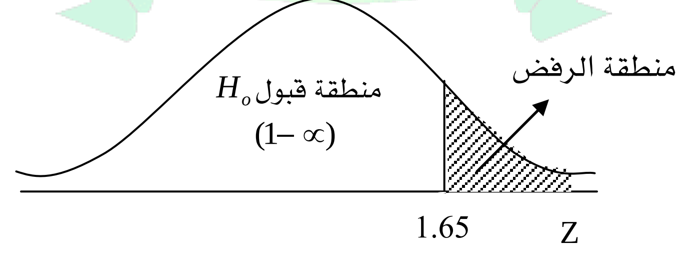
3. حساب إحصاءة الاختبار(قيمة Z المحسوبة):



$$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

$$Z = \frac{21-20}{10/\sqrt{100}} = \frac{1}{1} = 1$$

شكل(16.4)

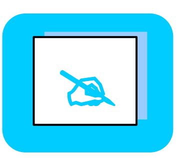


4. اتخاذ القرار:

بما أن قيمة Z المحسوبة(1) أقل من قيمة Z الجدولية 1.65 أي أنها تقع في منطقة قبول H_o أنظر شكل (16.4) .وعليه:

فإننا نقبل فرض العدم H_o عند مستوى معنوية $\alpha = 0.05$، ولا يوجد ما يدل على عدم صحة افتراض البنك من أن معدل فترة انتظار العميل لا تزيد على 20 دقيقة.

مثال (32.4):



إذا كان لدينا عينة وحداتها 4 , 5 , 5 , 3 , 2 مسحوبة من مجتمع طبيعي وسطه الحسابي $\mu = 2$ وانحرافه المعياري مجهول. فهل متوسط العينة يختلف معنوياً عن متوسط المجتمع عند مستوى معنوية 0.05

الحل:

أولاً: نقوم بحساب الوسط الحسابي:

$$\overline{X} = \frac{2+3+5+5+4}{5} = \frac{19}{5} = 3.1$$

ثم نحسب الانحراف المعياري

X_i	$(X_i - \overline{X})^2$
2	3.24
3	0.64
5	1.44
5	1.44
4	0.04
$\sum$	8.28

$$S = \sqrt{\frac{\sum (x_i - \overline{x})^2}{n}} = \sqrt{\frac{8.28}{5}}$$

$$= \sqrt{1.656} = 1.3$$

ثانياً:

1. صياغة الفرضيات.

$$\because H_o : \mu = 2$$

$$H_1 : \mu \neq 2$$

وحيث أن الفرضية البديلة H_1 تتضمن إشارة عدم التساوي $(\neq)$ فيكون الاختبار ذو جانبين.

2. تحديد مستوى المعنوية ومساحة مناطق القبول والرفض والنقاط الحرجة:

باستخدام الملحق رقم (2) وعند مستوى دلالة $0.025 = \frac{0.05}{2} = \frac{\alpha}{2}$ ودرجات حرية

$$n-1 = 5-1 = 4$$

فنجد أن قيمة t الجدولية هي:

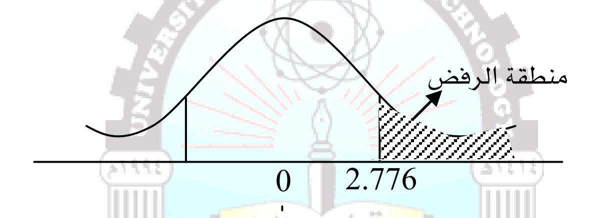
$$t_{1-\frac{\alpha}{2}\,,\,n-1} = t_{1-0.025,\,4} = t_{0.975,4} = 2.776$$

3. حساب إحصاءة الاختبار(قيمة t المحسوبة):

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

$$t = \frac{3.8-2}{1.3/\sqrt{5}} = \frac{1.8}{0.58} = 3.1$$

شكل(4-17)



4 . القرار: بما أن قيمة t المحسوبة أكبر من قيمة t الجدولية عند مستوى معنوية 0.025 ومع درجات حرية عددها 4 أي أنها تقع في منطقة رفض H_0 أنظر الشكل(17.4)، فإننا نرفض فرض العدم H_o ، ونقبل الفرض البديل أي أن هناك فرق معنوي بين متوسطي العينة والمجتمع.



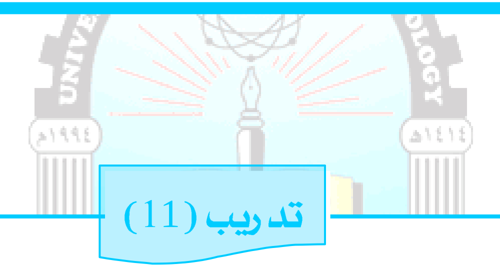
تدريب (9)

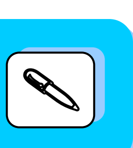
عرف الفرض الإحصائي. ثم أذكر الوظيفة الأساسية لاختبار الفروض الإحصائية في ظل عدم التأكد.



تدريب (10)

1. عرف فرض العدم (الفرضية الصفرية) وكذلك الفرض البديل
2. اذكر الخطوات الأساسية في اختبارات الفروض.





تدريب (11)

يعتقد أحد بائعي أجهزة الحاسوب أن متوسط ربحه في الجهاز الواحد يقل عن 500 ريال نتيجة لارتفاع نسبة الخصم المقدمة لعملائه. ولمعرفة مدى صحة هذا الادعاء ، سحبت عينة عشوائية من فواتير بيع سابقة فوجد أن متوسط الربح في العينة هو 481 ريالاً بانحراف معياري قدوه 47 ريالاً هل تؤيد هذه البيانات الفرض القائل بأن متوسط الربح يقل عن 500 ريال عند مستوى المعنوية $\alpha = 0.05$.



أسئلة التقويم الذاتي (4) :

1) ماذا يقصد بالاختيار من جهة واحدة والاختيار من جهتين ؟

2) يعتقد مدرس الاحتمالات أن متوسط درجات الطلاب في مقرر الاحتمالات الذي يقوم بتدريسه يزيد عن 74 ، لذلك قام بجمع بيانات لعينة من 34 درجة فوجد بأن الوسط الحسابي لهذه الدرجات يساوي 80 ، فإذا كان متغير الدرجات يتوزع طبيعياً بانحراف معياري يساوي 13. فهل تؤيد المدرس في اعتقاده عند مستوى معنوية 0.01.



نشاط

عزيزي الدارس : قم بزيارة لشركة الاتصالات اليمنية واحصل منهم على بيانات توضح مدى تطور الاتصالات في اليمن خلال العشرين السنة الماضية .. لكي تستخدمها في التنبؤ المستقبلي لليمن.



5. الخلاصة

ركزت هذه الوحدة على مفهوم الاحتمال والذي يعرف بأنه عبارة عن التكرار النسبي لحصول حدث معين عند إجرائه عدد كبير من المرات، وأن مجموع الاحتمالات يساوي واحد ، ولا يوجد احتمال سالب حيث أن قيمته تكون محصورة بين الصفر والواحد الصحيح.كما أن الحدث المؤكد هو الذي يساوي فضاء العينة.وأن أهم قوانين الاحتمالات هو قانون الجمع وقانون الضرب.

كما تناولت الوحدة المتغير العشوائي والذي يأخذ قيماً عددية تحددها نتائج التجربة العشوائية. وينقسم إلى قسمين رئيسين هما : المتغير العشوائي المنفصل وهو الذي يأخذ قيماً حقيقية صحيحة في مجال تغيره ، والمتغير العشوائي المتصل: بأنه المتغير الذي يمكن أن يأخذ عدداً غير منتهٍ من القيم في مجال تغيره. كما ركزت الوحدة على قانون التوزيع الاحتمالي من خلال استعراض واحد من أهم هذه التوزيعات الاحتمالية وهو التوزيع الطبيعي، والذي يعتبر حجر الأساس للنظرية الإحصائية الحديثة وهو أكثر التوزيعات الإحصائية المتصلة أهمية واستخداماً ويطلق عليه أحياناً توزيع جاوس. كما أن كثيراً من التوزيعات المهمة تُشتق منه أو تؤول إليه في حالات معينة وتأخذ دالة كثافة الاحتمال للتوزيع الطبيعي الشكل الآتية:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad , \quad \begin{array}{l} -\infty < x < +\infty \\ -\infty < \mu < +\infty \\ 0 < \sigma < +\infty \end{array}$$

ومن أهم خوص هذا التوزيع:

1- شكله متماثل ويشبه الجرس المقلوب

2. يمتد طرفا المنحنى من $-\infty$ إلى $+\infty$

3. المتوسطات تكون متساوية أي أن المتوسط الحسابي= الوسيط = المنوال

4. المجموع الكلي للمساحة تحت منحنى التوزيع الطبيعي تساوي الواحد الصحيح

5. تتوزع المساحة تحت المنحنى على المحور الأفقي كالأتي:

68٪ تقريباً من مساحة الشكل الموزع طبيعياً تقع بين $\mu \pm \sigma$

95٪ تقريباً من مساحة الشكل الموزع طبيعياً تقع بين $\mu \pm 2\sigma$

99.7٪ تقريباً من مساحة الشكل الموزع طبيعياً تقع بين $\mu \pm 3\sigma$

6.يعتمد شكل المنحنى على معلمتي الوسط الحسابي μ والانحراف المعياري σ

7. التوقع = μ ، التباين = σ^2 ، الانحراف المعياري = σ

وتحسب المساحة تحت المنحنى الطبيعي من تحويل قيم المتغير العشوائي المعتدل إلى قيم معيارية أي تحويله إلى متغير عشوائي جديد يسمى بالتوزيع الطبيعي القياسي والذي يرمز له بالرمز Z (وهو التوزيع الطبيعي المعياري الذي وسطه صفر وتباينه واحد) أي $Z = \frac{x-\mu}{\sigma}$

ثم تناولت الوحدة اختبارات الفروض الإحصائية والتي لها دور بالغ الأهمية في مساعدة الباحثين ، والإداريين من أجل ترشيد القرارات.

الفرضية الإحصائية: هي عبارة عن صياغة مبدئية حول معلمة أو أكثر من معالم المجتمع المجهولة. والوظيفة الأساسية لاختبار الفروض الإحصائية في ظل عدم التأكد من طبيعتها هي مساعدتنا في اتخاذ قرارات رشيدة بخصوص هذه المعالم.

فرض العدم: هو عبارة عن صياغة مبدئية عن معلمة المجتمع المجهولة ويرمز له بالرمز H_O.

الفرض البديل:هو عبارة عن صياغة مبدئية تشير إلى أن نفس المعلمة المجهولة لها قيمة تختلف عن القيمة التي حددها فرض العدم ويرمز له بالرمز H_1.

الخطوات الأساسية في اختبارات الفروض

1- التعرف على توزيع المجتمع الإحصائي

2- صياغة فرض العدم والفرض البديل

3- تحديد مستوى المعنوية .

4- صياغة قاعدة القرار.

5- اتخاذ القرار .

6. إجابات التدريبات :

تدريب(1):

1.أحداث مستحيلة: مثل ظهور الرقم 7 عند رمي حجر النرد مرة واحدة، أو ظهور الصورة والكتابة معاً عند رمي قطعة نقد مرة واحدة.
واحتمال وقوعه يساوي الصفر لأنه يحتوي على مجموعة خالية من العناصر.

2. **أحداث مستقلة**: مثل ظهور كرة حمراء عند سحب كرة بطريقة عشوائية مع الإرجاع من كيس يحتوي على كرات حمراء وصفراء .

3. **أحداث متنافية**: مثل حادثة ولادة طفل فإما أن يكون ذكراً و إما أنثى.وأيضاً حادثة نجاح الطالب في مقرر دراسي ينفي أو يمنع رسوبه في نفس المقرر.

تدريب(2):

نفرض أن :

A_1 : الورقة المسحوبة ثلاثة

A_2 : الورقة المسحوبة ثمانية

A_3 : الورقة المسحوبة صورة

أ. في حالة الحوادث المتنافية الوقوع

$$\therefore P(A_1 \cup A_2 \cup A_3) = P(A_1) + P(A_2) + P(A_3)$$

$$P(A_1 \cup A_2 \cup A_3) = \frac{4}{52} + \frac{4}{52} + \frac{12}{52} = \frac{5}{13}$$

تدريب(3):

نلاحظ أن حادثة الرسوب في الإحصاء لا تمنع أو تنفي الرسوب في المحاسبة. وبالتالي فالحادثتان مستقلتان وليستا متنافيتين. ونلاحظ أن الطالب الذي تم اختياره عشوائياً يمكن أن تنطبق عليه إحدى الحالات التالية:

راسب في الإحصاء

راسب في المحاسبة

راسب في الإحصاء والمحاسبة

وحيث أن السؤال قد حدد احتمال الرسوب فاحتمال النجاح مستبعداً.

فإذا رمزنا للحالة الأولى بالرمز $P(E_1)$ وللحالة الثانية بالرمز $P(E_2)$ ، فإن الرمز المناسب للحالة الثالثة هو $P(E_1 \cap E_2)$ وهو احتمال وقوع الحادثين معاً.

$$P(E_1) = \frac{18}{100} = 0.18 \quad , \quad P(E_2) = \frac{15}{100} = 0.15$$

$$P(E_1 \cap E_2) = \frac{8}{100} = 0.08$$

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$$

$$= 0.18 + 0.15 - 0.08 = 0.25$$

تدريب(4):

1. نعلم من مسلمات الاحتمال أن:

$$P(A) + P(A^c) = 1$$

وبالتعويض بالقيم الاحتمالية نجد أن:

$$P(A) + P(A^c) = 0.48 + 0.42 = 0.90 \neq 1$$

وهذا هو وجه الخطأ.

2. وجه الخطأ أن قيمة الاحتمال أكبر من الواحد الصحيح وقيمة الاحتمال دائماً محصورة بين الصفر والواحد.

3. وجه الخطأ أنه يتناقض مع مسلمة الاحتمال التي تقول: إن احتمال أي حادثة دائماً كمية موجبة.

تدريب(5):

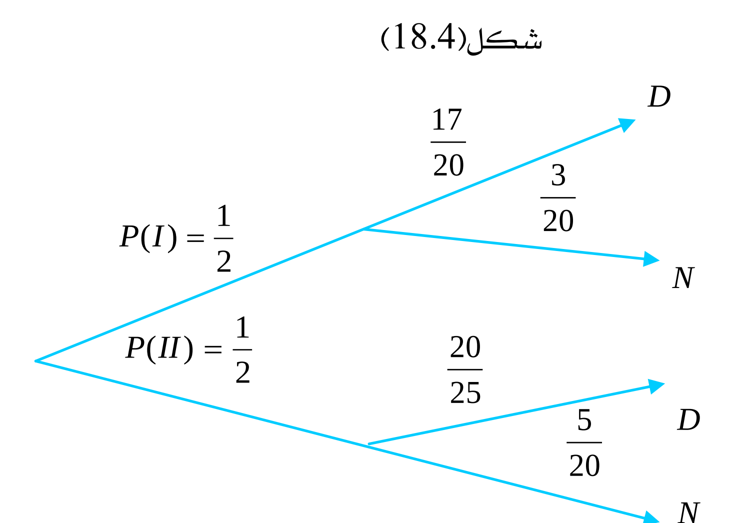
سوف نستعمل شجرة الاحتمال كإحدى الطرق المناسبة لحساب احتمال حادث معين إذا كان هذا الحادث يأتي من قبل حوادث متعددة تمثل تقسيماً لفضاء العينة ، حيث نمثل فضاء العينة بأصل الشجرة ونمثل تقسيم فضاء العينة بفروع الشجرة. ويستفاد من شجرة الاحتمال أيضاً في تمثيل نتائج التجربة الإحصائية المتعددة المراحل وتعيين احتمال كل فرع من فروع الشجرة التي مثلث تلك التجربة.

والآن إذا وضعنا الحرف D= عنصر تالف و N= عنصر صالح ثم نبني شجرة الاحتمال كما هو موضح في الشكل(18-4) أدناه مع بيانات احتمالات الفروع في المرحلة الأولى والاحتمالات الشرطية للفروع في المرحلة الثانية فإن:

(1) احتمال أن العنصر تالف هو مجموع احتمالات الحروف المنتهية بالحرف D

جامعة العلوم والتكنولوجيا

شكل(18.4)



$$P(D) = (\frac{1}{2})(\frac{17}{20}) + (\frac{1}{2})(\frac{20}{25}) = 0.825$$

(2) احتمال أن العنصر مسحوب من الصندوق الثاني علماً بأنه تالف هو :

$$P(II/D) = \frac{P(II)\,P(D/II)}{P(D)}$$

$$= \frac{\frac{1}{2} \cdot \frac{20}{25}}{0.825} = 0.48$$

تدريب(6):

1. المتغير العشوائي هو الذي يأخذ قيماً عددية تحددها نتائج تجربة عشوائية. وينقسم إلى قسمين رئيسيين هما: المتغير العشوائي المنفصل والمتغير العشوائي المتصل.

2. بعض خصائص منحنى التوزيع الطبيعي :

1- شكله متماثل ويشبه الجرس المقلوب مرتفع من الوسط ومنخفض بشكل تدريجي عند الطرفين ولكونه متماثل فإن 50٪ من مساحته تقع إلى يمين المتوسط الحسابي و 50٪ على يسار المتوسط الحسابي μ .

2. يأخذ الشكل الجرسي حيث يمتد طرفا المنحنى من $-\infty$ إلى $+\infty$ ويكون عدد التكرارات صغيرة عند طرفي المنحنى ولا يلتقيان بالمحور السيني أبداً وتزيد كلما اقتربت من مركز المنحنى.

3- إذا كان التوزيع معتدل فإن المتوسطات تكون متساوية أي أن المتوسط الحسابي= الوسيط = المنوال

4- المجموع الكلي للمساحة تحت منحنى التوزيع الطبيعي تساوي الواحد الصحيح.

5- يعتمد شكل المنحنى اعتماداً كاملاً على معلمتي الوسط الحسابي μ والانحراف المعياري σ

تدريب(7):

القيمة المعيارية Z تحسب من العلاقة الآتية: $Z = \frac{X - \mu_x}{\sigma_x}$

1. القيمة المعيارية لـ 700 ساعة

$$Z = \frac{700 - 650}{6} = 8.33$$

2. القيمة المعيارية لـ 690 ساعة

$$Z = \frac{690 - 650}{6} = 6.67$$

تدريب (8):

ليكن x هو متغير طول الأنبوب ولكي نحسب الاحتمال $P(X > 20)$ لابد لنا من المعايرة القياسية (حساب القيم المعيارية)كما سبق حيث سنحصل على

$$P(X > 20) = P\left(\frac{X - 24}{3} > \frac{20 - 24}{3}\right)$$

$$= P(Z > -1.33$$

وبهذه المعايرة يصبح المتغير $\frac{x - 24}{3}$ متغيراً قياسياً $X \sim N(0,1)$ أي أن هذه النقطة تبعد 1.33 انحراف معياري على يسار الوسط الحسابي وهي تحصر مساحة قيمتها

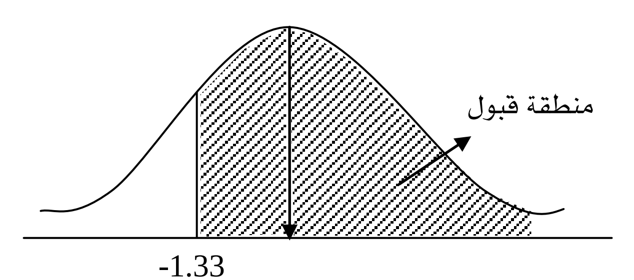
$$P(Z > -1.33) = 0.5 + P(0 \le Z \le 1.33)$$

$$= 0.5 + 0.4083 = 0.9083$$

انظر ملحق (1)

$$\therefore P(X > 20) \cong P(Z > -1.33) = 0.91$$

شكل(19.4)



ويعني هذا الاحتمال أن 91٪ فقط من الأنابيب يزيد طولها عن 20Cm وعليه فمنتج هذا المصنع لا يحقق هذا الطلب.

تدريب(9):

يعرف الفرض الإحصائي بأنه ادعاء أو افتراض حول متغير عشوائي أو متغيرات عشوائية. والوظيفة الأساسية لاختبار الفروض الإحصائية في ظل عدم التأكد من طبيعتها هي مساعدتنا في اتخاذ قرارات رشيدة بخصوص هذه المعالم . ونتيجة لذلك فإن الإحصاء الحديث يشار إليه غالباً بأنه (علم اتخاذ القرار في ظل عدم التأكد) واتخاذ القرار ما هو إلا اختيار من بين بدائل.

تدريب (10):

يعرف فرض العدم والذي يرمز له بـ H_o بأنه :" صياغة مبدئية عن معلمة المجتمع المجهولة " وتوحي هذه الصيغة إلى فكرة عدم وجود فرق بين معلمة المجتمع وقيمة معينة.

ويعرف الفرض البديل H_1 بأنه " صيغة مبدئية تشير إلى أن نفس المعلمة المجهولة لها قيمة تختلف عن القيمة التي حددها فرض العدم "

2. خطوات اختبار الفروض الإحصائية:

1- التعرف على توزيع المجتمع الإحصائي.

2- صياغة فرض العدم (null hypothesis)

والفرض البديل (alternative hypothesis)

3- تحديد مستوى المعنوية (level of Significance)

4- صياغة قاعدة القرار

5- اتخاذ القرار . وتنتهي هذه العملية باتخاذ قرار برفض أو عدم رفض فرض العدم.

تدريب(11):

نكتب فرض العدم والفرض البديل على الصورة:

$$H_0 : \mu = 500$$

$$H_1 : \mu < 500$$

ومن جدول t نلاحظ أن قيمة t الحرجة لاختبار ذو طرف أيسر عند $\alpha = 0.05$ ودرجة حرية 24 هي 1.711- حيث:

$$t_{1-\propto ,n-1} = t_{1-0.05\ 25-1}$$

$$= t_{0.95\ ,\ 24} = 1.711$$

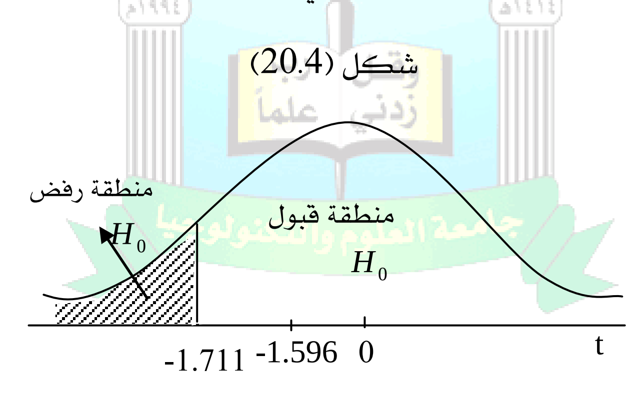
وهي قيمة سالبة لأن الفرض البديل ينص على أن متوسط الربح أقل من 500 ريال، أي أن الفرض البديل ذو طرف أيسر . وبالتالي تقع منطقة الرفض على يسار متوسط توزيع t وهو الصفر . وتكون قاعدة القرار هي:

رفض H_o عند ما تكون $t \le -1.711$ وتُحسب t من العلاقة:

$$t = \frac{485 - 500}{47/\sqrt{25}} = -1.596$$

وهي أكبر من قيمة t الحرجة وبالتالي لا يمكن رفض H_o عند $\alpha = 0.05$.

شكل (20.4)



7. قائمة المصطلحات :

- **أنظمة العد** Technique's of Counting : بعض الطرق لتحديد النتائج الممكنة لتجربة ما.

- **الاحتمالات** Probabilities : أحد فروع علم الرياضيات الهامة والتي اتسع نطاق تطبيقاتها حتى أصبح يشمل كافة العلوم الطبيعية والاجتماعية.

- **مسلمات الاحتمال** Probability Axiom : وهي حقائق أو أحكام نسلم بصحتها أو بمشروعيتها دون الحاجة إلى برهان.

- **الاحتمال الشرطي** Conditional Probability :هو الاحتمال الذي يتم حسابه لحادث ما بعد وقوع حادث آخر قبله.

- **المتغير العشوائي** Random Variable : المتغير العشوائي X على الفضاء العيني S هو دالة من المجموعة S إلى مجموعة الأعداد الحقيقية R . $X : S \rightarrow R$ بحيث تكون الصورة العكسية لأي فترة تنتمي إلى المدى حادثاً في S

- **التوزيع الطبيعي (جاوس)** The Normal Distribution (Gauss)

هو توزيع احتمالي لمتغير عشوائي متصل قيم المتغير فيه هي مدى المحور السيني كله، ويعد هذا التوزيع من أهم التوزيعات الاحتمالية، ويشبه الجرس المقلوب وهو متماثل حول الوسط الحسابي μ ومن أهم معالم التوزيع الوسط الحسابي والانحراف المعياري للمجتمع σ.

- **التوزيع الطبيعي القياسي** **Standard Normal Distribution** : هو توزيع احتمالي مشتق من التوزيع الطبيعي (حالة خاصة) بعد تحويل قيم التوزيع الطبيعي إلى قيم معيارية، وسطه الحسابي يساوي صفر وانحرافه المعياري واحد صحيح. ويتم حساب المساحات تحت منحناه باستخدام جداول خاصة.

- **اختبار الفروض** Hypothesis Testing :يعتبر اختبار الفرضيات أحد المواضيع الرئيسة للاستدلال الإحصائي ويهدف للوصول إلى قرار بشأن معلمة المجتمع من خلال قبول أو رفض فرض العدم. وإن اختبار الفرض هو قاعدة أو إجراء تؤدي إلى رفض أو عدم رفض فرض العدم.

الخطأ من النوع الأول والخطأ من النوع الثاني

error Type I error and Type II

يحدث الخطأ من النوع الأول عند رفض فرض العدم H_o بينما في الحقيقة يجب قبوله، واحتمال الوقوع في مثل هذا النوع من الخطأ يرمز له بالرمز α (ألفا) أما الخطأ من النوع الثاني فيقع في حالة قبولنا لفرضية العدم H_o بينما هو في الحقيقة خطأ، واحتمال الوقوع في مثل هذا النوع من الخطأ يرمز له بالرمز β (بيتا)

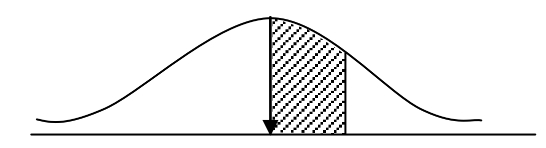
مستوى المعنوية Level of Significance

هي عبارة عن " احتمال رفض فرض العدم وهو صحيح " أي احتمال الوقوع في خطأ من النوع الأول Type 1 error ويرمز له بالرمز α.



8. الملاحق :

ملحق رقم (1) المساحة الواقعة تحت المنحنى الطبيعي القياسي



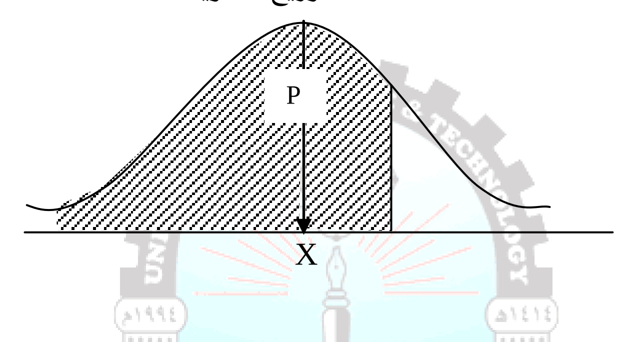
z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0639	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1062	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.7071	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2452	.2486	.2518	.2549
0.7	.2580	.2612	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.2718	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4083	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4682	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4485	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4773	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4865	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4965	.4957	.4945	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4980	.4980	.4981
2.9	.4981	.4982	.4983	.4983	.4984	.4989	.4985	.4985	.4986	.4986
3.0	.49865	.4987	.4987	.4988	.4988	.4998	.4989	.4990	.4990	.4990
3.1	.49903	.4991	.4911	.4991	.4992	.4992	.4992	.4992	.4993	.4993
3.2	.49931	.4993	.4994	.4994	.4994	.4994	.4994	.4995	.4995	.4995
3.3	.49952	.4995	.4995	.4996	.4996	.4996	.4996	.4996	.4496	.4497

3.4	.49966	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4998	.4998
3.5	.49977	.4998	.4998	.4998	.4998	.4998	.4998	.4998	.4998	.4998
3.6	.49984	.4998	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999
3.7	.49989	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999
3.8	.49993	.4999	.4999	.4999	.4999	.4999	.4999	.5000	.5000	.5000
3.9	.49995	.5000	.5000	.5000	.5000	.5000	.5000	.5000	.5000	.5000
4.0	.49997									

المصدر زياد رمضان صـ333.

ملحق رقم (2)

المساحات تحت توزيع استويدنت



V \ p	0.9	0.95	0.975	0.99	0.995
1	3.078	6.314	12.706	31.812	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921

17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763

المصدر البلداوي ملحق (6.6)



9. المراجع العربية والأجنبية:

المراجع العربية:

1. أبو صالح، محمد صبحي ومروة أحمد.(2005) : مبادئ الإحصاء. الطبعة الثانية، منشورات جامعة القدس المفتوحة، عمان: الأردن.
2. البلداوي، عبد الحميد .(1997) : الإحصاء للعلوم الإدارية والتطبيقية . الطبعة العربية الأولى، دار الشروق، عمان : الأردن.
3. حبيب، مجدي عبد الكريم .(2001): الإحصاء اللابارامتري الحديث في العلوم السلوكية. الطبعة الأولى. مكتبة النهضة المصرية، القاهرة : مصر
4. رمضان، زياد .(1997) : مبادئ الإحصاء الوصفي والتطبيقي والحيوي.الطبعة الرابعة، دار وائل للنشر، عمان : الأردن.
5. سمور، خالد قاسم وعبيد، أحمد جمعة .(1995): الاحتمالات . الطبعة الأولى، دار الفكر، عمان:الأردن.
6. عودة، أحمد. (1991) مقدمة في النظرية الإحصائية . الطبعة الأولى، منشورات جامعة الملك سعود.
7. القاضي، دلال وسهيلة والبياتي، محمود. (2004): الإحصاء للإداريين والاقتصاديين. دار الحامد للنشر والتوزيع، عمان: الأردن.
8. كنجو، أنيس إسماعيل .(2000) : الإحصاء والاحتمال . الطبعة الأولى، مكتبة العكيبان، الرياض : السعودية.
9. مصطفى، مدني دسوقي.(1968): مبادئ في نظرية الاحتمالات والإحصاء الرياضي وتطبيقاتها في الاستنتاج الإحصائي. دار النهضة العربية، القاهرة : مصر
10. المنصوب، محمد عبد الكريم .(1998): مفاهيم أساسية في الإحصاء . الطبعة الأولى، منشورات دار الخبرة، صنعاء: اليمن.
11. الهيتي، صلاح الدين حسين.(2004): الأساليب الإحصائية في العلوم الإدارية.الطبعة الأولى، دار وائل للطباعة والنشر، عمان: الأردن.

المراجع الأجنبية:

1- Cramer, H.(1961): Mathematical Methods of Statistics. Princeton: Princeton University press.

2- Feller, F.(1976): An Introduction to probability Theory and Its Applications, Vol . I, 3rd ed. New York: John Wiley.

3-Hayslett, H . T (J r), Statistics Made Simple, Made simple Books, Doubleday and Co ,.Inc Garden City , New York , 1968.

4- Hoel, P.; Port, S. and Stone , C. Introduction to Probability Theory . Boston: Houghton Miffin Company,1971.

5- Mendenhall , W . , and Beaver , R ., (1994) , Introduction to Probability and Statistics , 9th end , Duxbury Press.



8. التعيينات:

السؤال الأول : أختر الإجابة الصحيحة من بين الإجابات الواردة بوضع دائرة حول رقم الإجابة الصحيحة:

1- إن حوالي 95٪ من البيانات أو مساحة الشكل الموزع طبيعياً تقع على امتداد:

1- انحراف معياري واحد على كل جانب من القيمة μ

2- انحرافين معياريين على كل جانب من القيم μ

3- ثلاثة انحرافات معيارية على كل جانب من القيم μ

2- إذا كان A , B حادثتين متنافيتين فإن P(B\A) هو

(a) 1.0 (b) 0.0 (c) 0.5

3- عدد الطرق الكلية لترتيب حروف كلمة بلبل هي:

1) 5 , 2) 6 , 3) 4

4- صندوق يحوي كرتان زرقاء اللون و4 كرات بيضاء اللون

1) احتمال أن تكون الكرة المسحوبة في المرة الثانية زرقا اللون دون إعادة مع العلم بأن الأولى كانت زرقا هو :

1) $\frac{2}{6}$, 2) $\frac{1}{5}$, 3) $\frac{2}{5}$

2) احتمال أن تكون الكرة المسحوبة في المرة الثانية زرقا مع العلم بأن الكرة المسحوبة في المرة الأولى كانت بيضاء اللون هو :

1) $\frac{1}{5}$, 2) $\frac{2}{5}$, 3) $\frac{2}{6}$

5- إذا كان $A \subset B$ وكان $P(B)=0.7\ ,\ P(A)=0.4$ فإن

$P(B^c) = ...$ a) 0.6 , b) 0.3 c) 0.7

6- الحدث المؤكد هو عبارة عن:

1- مجموعة المشاهدات التي يمكن ظهورها عند إجراء تجربة ما

2- مجموعة جزيئه من فضاء العينة وقد يساويها

3- هو الحدث الذي يتألف من مجموعة خالية

7- في تجربة إلقاء حجري نرد مرة واحدة، وكان الحدث A يعنى الفرق المطلق بين العددين الظاهرين يساوي 2 فإن الاحتمال P هو:

1) 0.229　　2) 0.255　　3) 0.222

8- في تجربة إلقاء حجري نرد مرة واحدة فإن عدد عناصر فضاء العينة هو

1) 36　　, 2) 30　　, 3) 16

السؤال الثاني:

1- عرف كلاً من الاحتمالات الحسابية، فضاء العينة ، التجربة العشوائية .

2- ما الفرق بين الحوادث المتنافية والحوادث المستقلة مع التوضيح بأمثلة على كل منها.

3- كيس يحتوي على عشر كرات ثلاث منها حمراء والباقي من اللون الأبيض، سحبت كرة بشكل عشوائي من الكيس وأضيفت كرة من النوع المخالف للكرة المسحوبة . سحبت بعد ذلك كرة ثانية من الكيس.

1- أوجد احتمال P أن تكون الكرة الثانية حمراء

2- أوجد احتمال P أن تكون الكرتان من اللون الأبيض 3- أوجد احتمال P أن تكون الكرتان من نفس اللون

4- أوجد P(A/B) إذا كان 1- $A \subset B$ 2- A, B متنافيين

5- ليكن A , B حادثين بحيث أن

أوجد $P(A)=\frac{1}{2}$　, $P(B)=\frac{1}{3}$　, $P(A \cap B)=\frac{1}{4}$

$P(A/B)$ $2-P(B/A)$　$3-P(A \cup B)$　$4-P(A^c/B^c)$ $5-P(B^c/A^c)$

6- إذا كانت درجات حاصل الذكاء تتبع توزعاً طبيعياً بمتوسط حسابي يساوي 105 وانحراف معياري يساوي 15 ، فما نسبة الناس ذوي درجة الذكاء:

1- أكبر من 130 ؟

2- أقل من 85 ؟

3- بين 80 و 135 ؟

والله الموفق ،،،





ISBN 978-162154561-3
9 781621 545613

يطلب هذا الكتاب مباشرة من مركز جامعة العلوم والتكنولوجيا للكتاب الجامعي

Web Site: WWW. ust.edu/centers/ubc - Email: ubc@ust.edu - Tel: 00971 384078